# AN EVOLUTIONARY AND OPTIMISED APPROACH

# ON 3D-TV

C. Fehn[1], P. Kauff[1], M. Op de Beeck[2], F. Ernst[2], W. IJsselsteijn[3], M. Pollefeys[4], L. Van Gool[4], E. Ofek[5] and I. Sexton[6]

[1]Heinrich-Hertz-Institut, Germany ,[2]Philips Research, The Netherlands, [3]Eindhoven University of Technology, The Netherlands, [4]Katholike Universiteit Leuven, Belgium, [5]3DV Systems, Israel and [6]De Montfort University, United Kingdom

## ABSTRACT

In this paper we will present the concept of a system that allows for an evolutionary introduction of depth perception into the existing 2D digital TV framework. The work is part of the European Information Society Technologies (IST) project “Advanced Three-Dimensional Television System Technologies” (ATTEST), an activity where industries, research centers and universities have joined forces to design a backwards-compatible, flexible and modular broadcast 3D-TV system, where all parts of the 3D processing chain are optimised to one another. This includes content creation, coding and transmission, display and research in human 3D perception, which be will used to guide the development process. The goals of the project 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.

## INTRODUCTION

As early as in the 1920s, John Logie Baird, one of the TV pioneers, dreamed of developing high-quality, three-dimensional (3D) colour TV, as only such a TV system would provide the most natural viewing experience (2)(8). Today, eighty years later, the first black-and-white television prototypes have evolved into high-definition digital colour TV, but the hurdle of 3D still remains to be taken.

Discussions with major electronics companies reveal that 3D is expected to be the next major revolution in the history of TV. Both at professional and consumer electronics exhibitions, 3D-video and 3D displays 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.

In this paper, we describe 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 (1). This goal differs from that of previous 3D-TV projects (e.g. (3)(16)), which aimed primarily at technological progress.

In ATTEST, we will design an entire 3D-video chain, shown in Fig. 1, where all the different parts are optimised to one another. The need for 3D-video data will be satisfied by the development of a novel broadcast 3D camera that can be used to generate new content, and by converting existing 2D-video material into 3D. Coding schemes will be developed within the existing MPEG and DVB standards that allow for the transmission of depth information in one or more enhancement layers, ensuring full compatibility with existing decoders in 2D-TV set-top boxes (15). Furthermore, as at present, a suitable glasses-free 3D-TV display is not available, ATTEST will develop two 3D displays (single- and multiple user) that will allow free positioning. 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 iterated, user-centered design cycle.

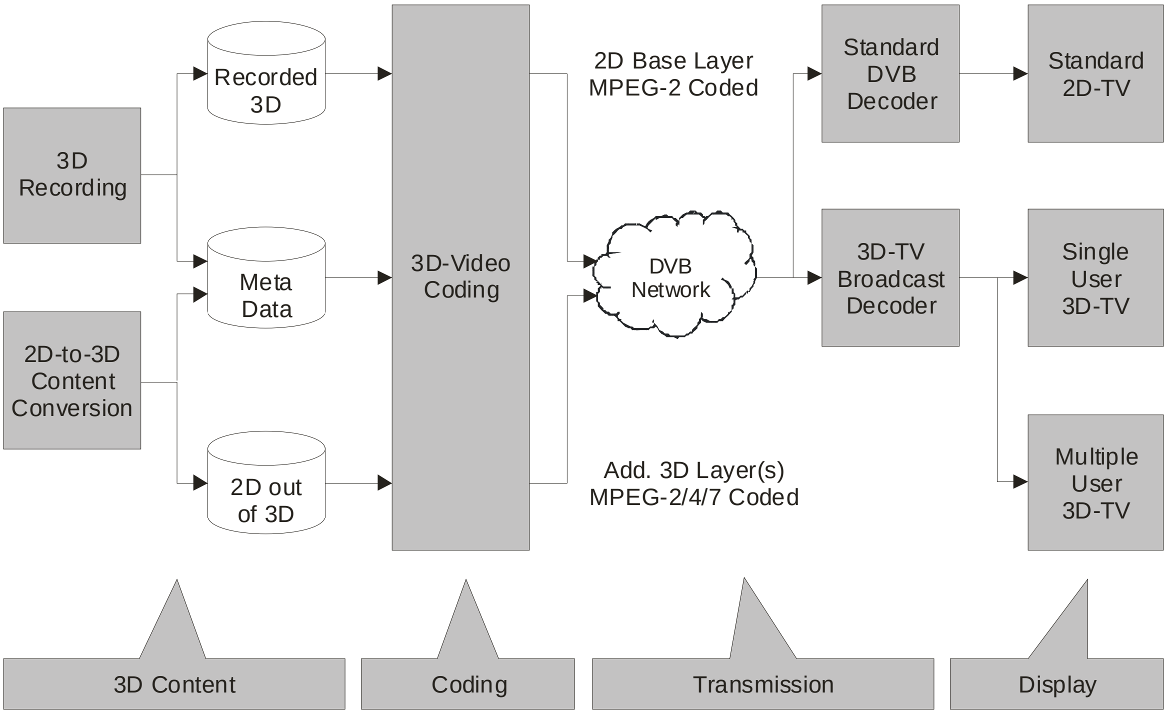



Figure 1 – The ATTEST 3D-video processing chain

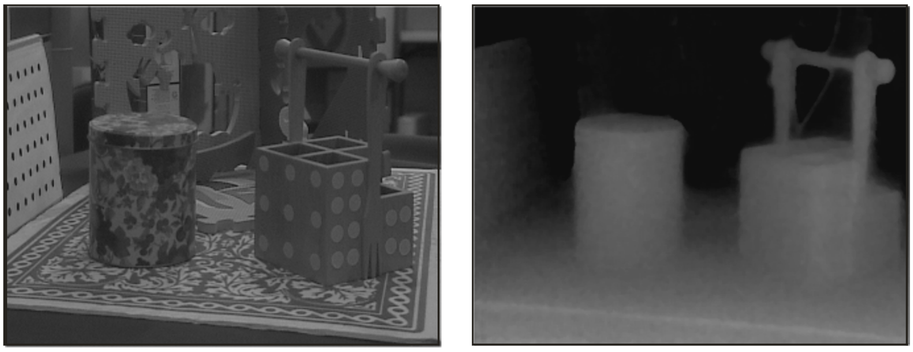
In the following sections we will elaborate in greater detail on the individual parts of the 3D-video processing chain. We will start with a brief description of the content generation part, where we will discuss the development of the novel 3D camera as well as the 2D-to-3D conversion. Special emphasis will then be put on the coding and transmission aspects. This will be followed by some results for the novel view generation, a short description of the 3D displays and the human perceptual evaluation of the 3D-video chain. Finally, we will make an analogy to the introduction of the previous revolutionary feature in TV: colour. This will help to elucidate some of the decisions we made during the set-up of the project. We will finish the paper with a short conclusion.

## CONTENT GENERATION

It is clear, that the success of any future 3D-TV broadcast system will depend to a great extent on the timely availability of sufficient exciting and/or interesting 3D-video material. Therefore, methods for the fast and cheap generation of suitable 3D-video content must be developed. In ATTEST, the need for 3D-video data will be satisfied in two different ways:

### Novel 3D-video camera

For the generation of new 3D-video content, ATTEST will convert an existing IR range camera, the so-called Zcam$^{TM}$ from 3DV Systems, into a full 3D camera. The Zcam$^{TM}$ works like an infrared LIDAR (laser radar without mechanical scanner) system integrated into a traditional video capture device, and is designed as an add-on to an existing broadcast camera (9). ATTEST will improve the original Zcam$^{TM}$ to deliver a high depth and pixel resolution 3D camera, fit for indoor use. The output of the ATTEST 3D camera will consist of two separate video streams: broadcast quality RGB and associated depth with pixel resolution accuracy. An example of a video- and a depth image captured with the current Zcam$^{TM}$ is shown in Fig. 2.



(a) (b)

Figure 2 – Images taken by the depth camera; (a) normal RGB image; (b) accompanying depth image (graylevel inversely proportional to depth)

### Conversion of existing 2D content

It seems clear, that the need for 3D-video content can only partially be satisfied with newly recorded material. Therefore, ATTEST will also develop depth reconstruction algorithms that can be used to convert existing 2D-video material into 3D. In contrast to the techniques described in (7), ATTEST aims at tools that only require minimal manual interaction:

- Accurate off-line conversion tools will be provided for use at the broadcaster and content provider side. These will primarily be used to convert popular movies and impressive documentaries. As there aren't any real-time constraints in this case, all available video data can be used in the computations and 3D information can be integrated over a whole shoot, resulting in high quality 3D reconstructions. Therefore, the ATTEST approach will build further upon techniques developed for 3D modelling from image sequences (17).
- For the consumer side, ATTEST will develop on-line conversion methods (e.g. for processing in a set-top box), that will allow the viewer to augment any suitable, incoming 2D broadcast to 3D. In this case, computations can only be based on video frames that have already been received. Therefore, it is planed to further extent the techniques described in (19).

## 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 which often relied on the basic concept of 'stereovision', 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 (5)(6). For this purpose, the data representation and coding syntax of the ATTEST system make use of the layered structure shown in Fig. 3. This structure basically 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 layer(s) deliver(s) the additional information to the 3D-TV receiver. The minimum information transmitted in the enhancement layer(s) 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 occlusions) it might be useful to send further information, for example segmentation masks and occluded texture. Note that the layered structure in Fig. 3 is extendable in this sense.

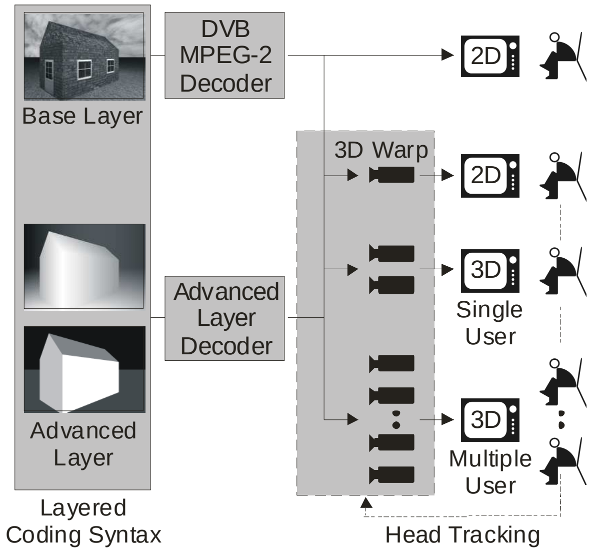



Figure 3 – Layered coding syntax

For the transmission of the enhancement layer(s), it is planed to rely as far as possible on already available MPEG-2/4/7 tools. Nevertheless, in case that existing tools prove to be inadequate to meet the ATTEST project goals, we are intending to contribute our results to the appropriate standardisation bodies. Therefore, ATTEST is already strongly participating in the recently established MPEG Ad-hoc group (AHG) on 3D-video coding (14).

Additionally it is important to realise that stereovision is only one of the relevant depth cues and that other cues such as motion-parallax, texture, brightness and geometric appearance of video objects are of comparable importance (4). For scene objects that are sufficiently far away from the viewer, they can even become dominant (see Fig. 4). It is therefore a significant feature of the described layered structure that it is flexible enough to support alternative forms of depth representation. This allows for a 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 watcher (see also Fig. 3). By this means, broadcasters could provide the user with a first, limited depth impression through parallax viewing, even on conventional 2D-TV screens. On the other hand side, users willing to invest into a 3D-TV set could enjoy a full-blown stereo reproduction of the same data on single- or even multiple user 3D displays (see Fig. 3).

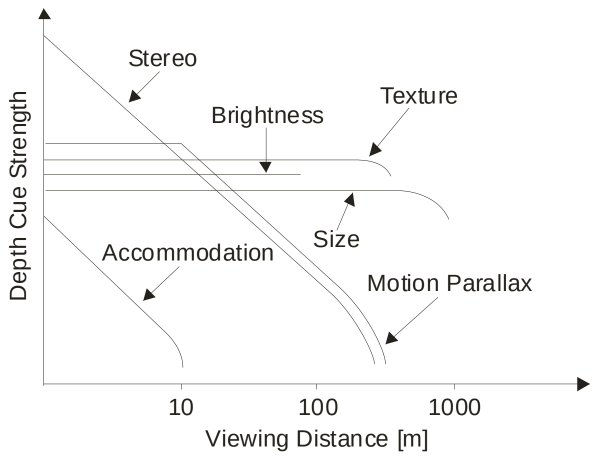


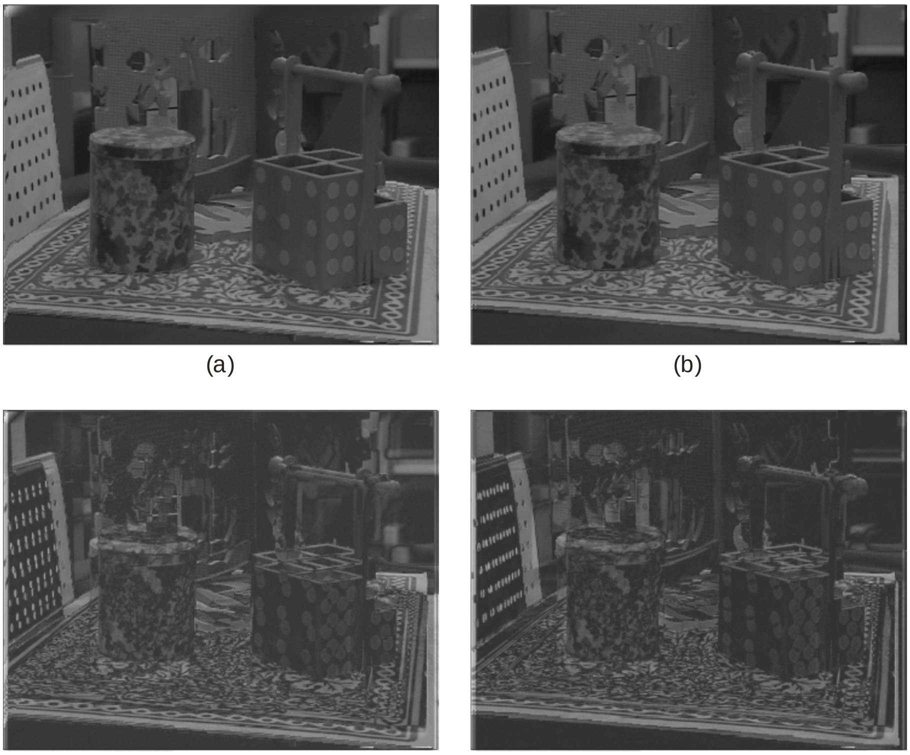

Figure 4 – Importance of depth cues in dependence of the viewing distance (figure after (12))

Another important 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 (13). Hence in our view, the TV watcher 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 could 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.

## NOVEL VIEW GENERATION

A key feature of the modular 3D-video chain will be the flexibility towards the consumer's display, which can be a conventional 2D display, or a single- or even a multiple user 3D screen. For the 2D display, the depth information can either be neglected, or used to render individual views based on the user's head-tracked viewing position (motion parallax). In the later case and for any kind of 3D display, one or more novel views with the correct perspective have to be synthesised from the transmitted video and depth information.

This can be done by using so-called image-based rendering (IBR) methods, which use the per-pixel depth information to warp the original image points into the desired novel view. An example that utilises an IBR method similar to the one described in (5) is shown in Fig. 5 (a-b). Here we synthesised two different novel views around the original viewing position shown in Fig. 2. The shift in perspective from the original- to the novel views is illustrated in Fig. 5 (c-d), where we show the per-pixel difference between the respective images.



(a) (b) (c) (d)

Figure 5 – Novel views created by using IBR methods; (a-b) Novel views; (c-d) Per-pixel difference between the original- and the respective novel view

## 3D DISPLAYS

As already described in the last section, the whole ATTEST 3D-video processing chains is designed to be flexible towards the use of different 3D displays. Here we see 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 single user autostereoscopic display, based on a 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 be positioned close to an optimum viewing plane, as in the case of most currently available systems. Within the viewing region, the viewer will be provided with a picture with excellent depth quality, good colour reproduction and very low crosstalk. Hardware interfaces will enable live video to be displayed in real-time.
- The multi-user autostereoscopic display provides 3D-video for up to four viewers who can occupy a viewing area that is between one and three meters from the screen and ± 30° from the axis. Regions where the left image only, and the right image only, are seen across the complete width of the screen are referred to as the exit pupils. The multi-user display operates by having the conventional LCD backlight replaced by steering optics that are able to form the multiple steerable exit pupils, which utilise a combination of white LCD arrays, a two-dimensional spatial light modulator and a novel optical configuration to control the light direction. Head tracking is employed to ensure that these pupils are always located at the appropriate eyes.

## PERCEPTUAL EVALUATION

As we position the user experience at the focal point of the project, human perception studies will play a prominent role in ATTEST. Therefore, the project is set up in a way, that all individual steps in the 3D-video processing chain can be evaluated separately. The goal is to arrive at a set of requirements and recommendations for an optimal 3D-TV system and to contribute to each individual step in the chain through a perceptual and usability analysis 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 (10)(11)(18). 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.

## ANALOGY TO INTRODUCTION OF COLOUR

In the following, we will make an short analogy between the introduction of 3D-TV and the introduction of the previous revolutionary feature in TV: colour. This will help to elucidate some of the decisions we made during the set-up of the project:

- When colour was introduced to broadcast TV, only a few, newly recorded programs were transmitted in colour. Existing content could only be viewed in black-and-white, as re-colouring of monochrome movies only became mature very late in the transition period from monochrome to colour TV. While at that time, the lack of colour content was considered to be acceptable, the consumer's expectation towards novel technology has changed. To prevent similar problems during the introduction of 3D-TV, ATTEST puts strong emphasis on the development of suitable 3D reconstruction techniques, that will enable broadcasters and content providers to easily create sufficient 3D-video material from original 2D footage.

- The transmission of the analog colour TV was designed such that owners of black-and-white TV sets could still watch black-and-white video with unchanged quality. In a similar way, ATTEST aims at a transmission of 3D-video such that owners of a digital 2D TV set can watch the 3D digital TV broadcasts in 2D without a degradation in visual quality.
- In an analog colour TV system, the colour experience can be controlled by the viewer through the gearing of hue and saturation. This allows the user to adapt the colour reproduction to his own personal demand. Furthermore, a desaturation can also be used as a kind of graceful degradation in the case of unwanted transmission noise. The layered coding syntax used in ATTEST is designed to give the 3D-TV user a similar kind of freedom over the adjustment of the depth reproduction.

## CONCLUSION

We described 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 will cooperate towards a flexible, 2D-compatible and commercially feasible 3D-TV system for broadcast environments.

The ATTEST 3D-TV system will be comprised of an entire 3D-video processing chain, including content creation, coding, transmission and display. All parts of the chain will be optimised in an iterative user-centered design cycle, guided by research on human 3D perception.

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.

## 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), Piotr Wilinski 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.

## REFERENCES

1. ATTEST, 2002 – 2004. European IST-2001-34396 project. http://www.extra.research.philips.com/euprojects/attest/. 2002 – 2004.
2. Baird, M. H. I., 1996. Eye of the World: John Logie Baird and Television (Part II). http://www.arts.uwaterloo.ca/fine/juhde/hills961.htm. 1996.
3. DISTIMA, 1992 – 1995. European Race 2045 project. http://www.tnt.uni-hannover.de/project/eu/distima. 1992 – 1995.

4. Cutting, J. E., Vishton, P. M., 1995. Perceiving Layout and Knowing Distances: The Interaction, Relative Potency, and Contextual Use of Different Information about Depth. In Epstein, W. and Rogers, S. (Eds.), Handbook of Perception and Cognition: Perception of Space and Motion. Vol. 5. Academic Press. 1995. pp. 69 to 117.

5. Fehn. C., Cooke, E., Schreer, O. and Kauff, P., 2002. 3D Analysis and Image-Based Rendering for Immersive TV Applications. To appear in Special Issue on Image Processing Techniques for Virtual Environments and 3D Imaging of the Signal Processing: Image Communication Journal. October 2002.

6. Fehn, C. and Kauff, P., 2002. Interactive Virtual View Video (IVVV) – The Bridge Between 3D-TV and Immersive TV. To appear in Proceedings of SPIE IT105: Three-Dimensional TV, Video and Display. July – August 2002.

7. Harman, P., 2000. Home Based 3D Entertainment – An Overview. Proceedings of International Conference on Image Processing. September 2000.

8. Hills, A. R., 1996. Eye of the World: John Logie Baird and Television (Part I). http://www.arts.uwaterloo.ca/fine/juhde/baird962.htm. 1996.

9. Iddan, G. J. and Yahav. G., 1994. 3D Imaging in the Studio (and Elsewhere …). Proceedings of SPIE 4298: Videometrics and Optical Methods for 3D Shape Measurements. January 2001. pp. 48 to 55.

10. IJsselsteijn, W. A., de Ridder, H. and Vliegen J., 2000. Subjective Evaluation of Stereoscopic Images: Effects of Camera Parameters and Display Duration. IEEE Transactions on Circuits and Systems for Video Technology. Vol. 10. Nr. 2. March 2000. pp. 225 to 233.

11. IJsselsteijn, W. A., Freeman, J., Bouwhuis, D. G. and de Ridder, H., 2002. Presence as an Experimental Metric for 3D Display Evaluation. To be presented at the International Symposium of the Society for Information Display. May 2002.

12. Nagata, S., 1984. How to Reinforce Perception of Depth in Single Two-Dimensional Pictures. Proceedings of the SID. Vol . 25. Nr. 3. 1984. pp. 239 to 246.

13. Norman, J. F., Dawson, T. E., Butler, A. K., 2000. The Effects of Age Upon the Perception of Depth and 3-D Shape from Differential Motion and Binocular Disparity. Perception 2000. Vol. 29. Nr. 11. November 2000. pp. 1335 to 1359.

14. 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.

15. Op de Beeck, M. and Redert, A., 2001. Three Dimensional Video for the Home. Proceedings of International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging. May – June 2001. pp. 188 to 191.

16. PANORAMA, 1995 – 1998. European ACTS AC092 project. http://www.tnt.uni-hannover.de/project/eu/panorama. 1995 – 1998.

17. Pollefeys, M., Koch. R., Vergauwen, M. and Van Gool, L., 1999. Hand-Held Acquisition of 3D Models with a Video Camera. Proceedings of Second International Conference on 3-D Digital Imaging and Modelling. October 1999. pp. 14 to 23.

18. Stelmach, L., Tam, W. J. and Meegan D., 1999. Perceptual Basis of Stereoscopic Video. Proceedings of the SPIE 3639: Stereoscopic Displays and Virtual Reality Systems VI. January 1999. pp. 260 to 265.

19. Wilinski. O. and van Overveld. K., 1998. Depth From Motion Using Confidence Based Block Matching. Proceedings of Image and Multidimensional Signal Processing Workshop. July 1998. pp. 159 to 192.