Available online at www.sciencedirect.com
ScienceDirect
Signal Processing: *Image Communication* 22 (2007) 217–234
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

# Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability

P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer*, A. Smolic, R. Tanger

*Fraunhofer Institute for Telecommunications/Heinrich-Hertz-Institut, Einsteinufer 37, 10587 Berlin, Germany*

Received 17 November 2006; accepted 29 November 2006

**Abstract**

Due to enormous progress in the areas of auto-stereoscopic 3D displays, digital video broadcast and computer vision algorithms, 3D television (3DTV) has reached a high technical maturity and many people now believe in its readiness for marketing. Experimental prototypes of entire 3DTV processing chains have been demonstrated successfully during the last few years, and the motion picture experts group (MPEG) of ISO/IEC has launched related ad hoc groups and standardization efforts envisaging the emerging market segment of 3DTV. In this context the paper discusses an advanced approach for a 3DTV service, which is based on the concept of video-plus-depth data representations. It particularly considers aspects of interoperability and multi-view adaptation for the case that different multi-baseline geometries are used for multi-view capturing and 3D display. Furthermore it presents algorithmic solutions for the creation of depth maps and depth image-based rendering related to this framework of multi-view adaptation. In contrast to other proposals, which are more focused on specialized configurations, the underlying approach provides a modular and flexible system architecture supporting a wide range of multi-view structures.


*Keywords:* 3DTV; Interoperability; Multi-view adaptation; Disparity matching; Multi-baseline stereo; Depth image-based rendering; Multi-view coding; Auxiliary video data representation

## 1. Introduction

The history of 3D movies is as long as the story of the motion picture itself. Based on the technical concept of the Wheatstone stereoscope, which was a very popular entertainment device at the end of the 19th century, the Lumière brothers were the first to show moving 3D pictures at the 1903's world fair in Paris. The first full-length 3D movie was then presented in Los Angeles in 1922, and, in 1928, John Logie Baird applied the principle of stereoscopy to an experimental 3D television (3DTV) set-up using Nipkow's perforated disc technology [3].

In spite of these early demonstrations and many subsequent efforts to establish 3D movies, especially in cinema of the 1950s, the commercial breakthrough failed to appear. Technical deficiencies and insufficient quality hampered a successful introduction in these market segments. In fact, convincing 3D applications could only be found outside of cinema and TV for a long time, particularly in niche

*Corresponding author. Tel.: +49 30 31002 620.
*E-mail addresses:* Peter.Kauff@fraunhofer.hhi.de (P. Kauff), oliver.schreer@hhi.fraunhofer.de (O. Schreer).
*URL:* http://ip.hhi.de.

0923-5965/$ - see front matter 
doi:10.1016/j.image.2006.11.013

markets like simulators and visualization devices. The only remarkable exception in the entertainment segment is IMAX 3D, which has some commercial success since its introduction in 1986, and which launches a limited number of specialized 3D productions each year. The situation gradually changed in the early 1990s with the foreseeable transition from analogue to digital TV services. Worldwide research activities were started with the aim to develop standards, technologies and production facilities for 3DTV. The multi-view profile (MVP), which had been approved as part of the motion picture experts group's MPEG-2 standard in 1996, is one result of this lively research period [18].

However, all these 3DTV proposals from the early 1990s relied on the straightforward concept of an end-to-end stereoscopic video chain, i.e., on the capturing, transmission and display of two separate video streams, one for the left and one for the right eye. Due to these restrictions, stereo capturing had to fit to the display geometry and vice versa. Display properties and viewing conditions as well as related human factor aspects [17] had to be taken into account by the camera man during shooting like Panum's fusional horopter range within which the human visual system is able to merge two separately displayed views into one single stereo percept. Clearly, these constraints made 3D productions extremely complicated.

In the late 1990s researchers therefore came to the conclusion that it is essential to separate capturing and display geometry by using emerging methods of computer vision, 3D video processing and image-based rendering (IBR). The main idea was to derive an almost generic depth-based data representation from captured images to decouple camera and display geometry and, with it, to obtain highest flexibility and adaptability at the display side. On principle, this can be achieved by estimating depth information from a given stereo or multi-view camera system and to use these depth data to recalculate at the receiver side a virtual stereo pair, which is perfectly adapted to display properties and related viewing conditions. In 1998, the European PANORAMA project was one of the first research activities that could demonstrate the feasibility and the potential of such a depth-based process of stereo adaptation [27].

While the PANORAMA project was focused on stereoscopic videoconferencing, the European follow-up project ATTEST has taken up this concept later on and has applied it to the requirements of a 3DTV processing chain. The ATTEST system is based on the transmission of regular video images enriched with depth maps providing a $Z$-value for each pixel—a data representation, which is often called video-plus-depth therefore (see Fig. 1). The final stereo images are then reconstructed at the receiver side by using depth image-based rendering (DIBR) techniques. With it, the ATTEST concept has some crucial advantages over former 3DTV proposals, such as backwards compatibility to existing 2D services for digital video broadcast (DVB), efficient compression capabilities and a high adaptability to 3D display properties, viewing conditions, and user preferences [9].

Together with the advent of DVB and the enormous progress of 3DTV-related technology, particularly in the area of auto-stereoscopic 3D displays, these latest research results now seem to form a critical mass for the successful introduction of new, innovative entertainment services. It is widely accepted that the video-plus-depth structure from Fig. 1 provides the groundwork for near-future 3DTV systems, especially for two reasons. It

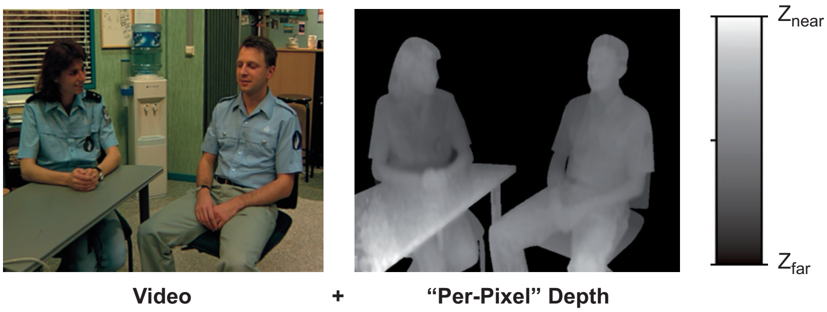


Fig. 1. The 3DTV data representation using video-plus-depth.