

Three-dimensional Television: From Science-fiction to Reality

Levent Onural and Haldun M. Ozaktas

Department of Electrical Engineering, Bilkent University, TR-06800 Bilkent,
Ankara, Turkey

Moving three-dimensional images have been depicted in many science-fiction films. This has contributed to 3D video and 3D television (3DTV) to be perceived as ultimate goals in imaging and television technology. This vision of 3DTV involves a ghost-like, yet high quality optical replica of an object that is visually indistinguishable from the original (except perhaps in size). These moving video images would be floating in space or standing on a tabletop-like display, and viewers would be able to peek or walk around the images to see them from different angles or maybe even from behind (Fig. 1.1). As such, this vision of 3DTV is quite distinct from stereoscopic 3D imaging and cinema.

3D photography, cinema, and TV actually have a long history; in fact, stereoscopic 3D versions of these common visual media are almost as old as their 2D counterparts. Stereoscopic 3D photography was invented as early as 1839. The first examples of 3D cinema were available in the early 1900s. Various forms of early 2D television were developed in the 1920s and by 1929, stereoscopic 3DTV was demonstrated.

However, while the 2D versions of photography, cinema, and TV have flourished to become important features of twentieth century culture, their 3D counterparts have almost disappeared since their peak around 1950. Our position is that this was not a failure of 3D in itself, but a failure of the then only viable technology to produce 3D, namely stereoscopy (or stereography). Stereoscopic 3D video is primarily based on the binocular nature of human perception, and it is relatively easy to realize. Two simultaneous conventional 2D video streams are produced by a pair of cameras mimicking the two human eyes, which see the environment from two slightly different angles. Then, one of these streams is shown to the left eye, and the other one to the right eye. Common means of separating the right-eye and left-eye views are glasses with colored transparencies or polarization filters. Although the technology is quite simple, the necessity to wear glasses while viewing has often been considered as a major obstacle in front of wide acceptance of 3DTV. But perhaps more importantly, within minutes after the onset of viewing, stereoscopy frequently causes eye fatigue and feelings similar to that experienced during motion



Fig. 1.1. Artist's vision of three-dimensional television. (graphic artist: Erdem Yücel)

sickness, caused by a mismatch of perceptory cues received by the brain from different sensory sources. Recently, with the adoption of digital technologies in all aspects of motion picture production, it has become possible to eliminate some of the factors which result in eye fatigue. This development alone makes it quite probable for stereoscopic 3D movies to be commonplace within a matter of years. Nevertheless, some intrinsic causes of fatigue may still remain as long as stereoscopy remains the underlying 3D technology.

Stereoscopic 3D displays are similar to conventional 2D displays: a vertical screen or a monitor produces the two video channels simultaneously, and special glasses are used to direct one to the left eye and the other to the right eye. In contrast, autostereoscopic monitors are novel display devices where no special glasses are required. Covering the surface of a regular high-resolution digital video display device with a vertical or slanted lenticular sheet, and driving these monitors by so-called interzigged video, one can deliver the two different scenes to the left and the right eyes of the viewer, provided that the viewer stays in the correct position. (A lenticular sheet is essentially a transparent film or sheet of plastic with a fine array of cylindrical lenses. The ruling of the lenses can either be aligned or slanted with respect to the axes of the display.) Barrier technology is another way of achieving autostereoscopy: electronically generated fence-like optical barriers coupled with properly interzigged digital pictures generate the two or more different views required. It is possible to provide many more views than the two views of classical stereoscopy, by using the autostereoscopic approach in conjunction with slanted lenticular sheets or barrier technology. Up to nine views are common, creating

horizontal parallax with a viewing angle of about 20 degrees. Classical stereoscopy with its two views is not able to yield parallax in response to head movement. People watching three-dimensional scenes expect occlusion and disocclusion effects when they move with respect to the scene; certain parts of objects should appear and disappear as one moves around. This is not possible with two fixed views, producing an unnatural result if the observer is moving. Head-tracking autostereoscopic display devices have been developed to avoid this viewer position constraint; however, serving many users at the same time remains a challenge.

Free viewpoint video (FVV) functionality is another approach to allowing viewer movement. It offers the same functionality familiar from three-dimensional computer graphics. The user can choose a viewpoint and viewing direction within a visual scene interactively. In contrast to pure computer graphics applications which deal with synthetic images, FVV deals with real world scenes captured by real cameras. As in computer graphics, FVV relies on a certain three-dimensional representation of the scene. If from that three-dimensional representation, a virtual view (not an available camera view), corresponding to an arbitrary viewpoint and viewing direction can be rendered, free viewpoint video functionality will have been achieved. In most cases, it will be necessary to restrict to some practical limits the navigation range (the allowed virtual viewpoints and viewing directions). Rendering stereo pairs from the three-dimensional representation not only provides three-dimensional perception, but also supports natural head-motion parallax.

Despite its drawbacks, stereoscopic 3D has found acceptance in some niche markets such as computer games. Graphics drivers that produce stereo video output are freely available. With the use of (very affordable) special glasses, ordinary personal computers can be converted into three-dimensional display systems, allowing three-dimensional games to be played. Stereo video content is also becoming available. Such content is either originally captured in stereo (such as in some commercially available movies) or is converted from ordinary two-dimensional video. Two-dimensional to three-dimensional conversion is possible with user-assisted production systems, and is of great interest for content owners and producers.

Stereoscopic 3D, whether in its conventional form as in the old stereoscopic cinema, or in its more modern forms involving autostereoscopic systems, falls far from the vision of true optical replicas that have been outlined at the beginning of this chapter. To circumvent the many problems and shortcomings of stereoscopy in a radical manner, it seems necessary to abandon the basic binocular basis of stereoscopy, and by turning to basic physical principles, to focus on the goal of true optical reconstruction of optical wave fields. Optically sensitive devices, including cameras and human eyes, do not “reach out” to the environment or the objects in them; they merely register the light incident on them. The light registered by our eyes, which carries the information about the scene, is processed by our visual system and brain, and thus we perceive our environment. Therefore, if the light field which fills a given 3D region can be

recorded with all its physical attributes, and then recreated from the recording in the absence of the original object or scene, any optical device or our eyes embedded in this recreated light field will “see” the original scene, since the light incident on the device or our eyes will be essentially indistinguishable in both cases. This is the basic principle of holography, which is a technique known since 1948. Holography is distinct from ordinary photography in that it involves recording the entire optical field, with all its attributes, rather than merely its intensity or projection (“holo” in holography refers to recording of the “whole” field). As expected, the quality of the holographic recording and reconstruction process will directly affect the fidelity of the created ghost-like images to their originals. Digital holography and holographic cinema and TV are still in their infancy. However, advances in optical technology and computing power have brought us to the point where we can seriously consider making this technology a reality. It seems highly likely that high quality 3D viewing will be possible as the underlying optics and electronics technologies mature.

Integral imaging (or integral photography) is an incoherent 3D photographic technique which has been known since 1905. In retrospect, the technique of integral imaging can also be classified as a kind of holography, since this technique also aims to record and reproduce the physical light distribution. The basic principle is to record the incidence angle distribution of the incoming light at every point of recording, and then regenerate the same angular illumination distribution by proper back projection. The same effect is achieved in conventional holography by recording the phase and amplitude information simultaneously, instead of the intensity-only recording of conventional photography. The phase information is recorded using interference, and therefore, holographic recordings require coherent light (lasers). Intensity recording, such as with common optical emulsion or digital photography, loses the direction information.

It is helpful to keep in mind the distinction between 3D displays and 3D television (3DTV). We use the term 3D display to refer to imaging devices which create 3D perception as their output. 3DTV refers to the whole chain of 3D image acquisition, encoding, transport/broadcasting, reception, as well as display. We have so far mostly discussed the display end of 3DTV technology. An end-to-end 3DTV system requires not only display, but also capture and transmission of the 3D content. Some means of 3D capture were already implicit in our discussion of displays. For example, stereoscopic 3DTV involves a stereoscopic camera, which is nothing but two cameras rigidly mounted side by side with appropriate separation. The recording process in integral imaging is achieved using microlens arrays, whereas holographic recording employs coherent light and is based on optical interference. In these conventional approaches, the modality of 3D image capture is directly related to that of 3D image reconstruction, with the reconstruction process essentially amounting to reversal of the capture process. In contrast, current research in 3DTV is targeting a quite different approach in which the input capture and output display

modalities are completely decoupled and bridged by digital representation and processing.

In recent years, tremendous efforts has been invested worldwide to develop convincing 3DTV systems, algorithms, and applications. This includes improvements over the whole processing chain, including image acquisition, three-dimensional representation, compression, transmission, signal processing, interactive rendering, and display (Fig. 1.2). The overall design has to take into account the strong interrelations between the various subsystems. For instance, an interactive display that requires random access to three-dimensional data will affect the performance of a coding scheme that is based on data prediction.

The choice of a certain three-dimensional scene representation format is of central importance for the design of any 3DTV system. On the one hand, it sets the requirements for acquisition and signal processing. On the other hand, it determines the rendering algorithms, degree of and mode of interactivity, as well as the need for and means of compression and transmission. Various three-dimensional scene representations are already known from computer graphics and may be adapted to 3DTV systems as well. These include different types of data representations, such as three-dimensional mesh models, multiview video, per-pixel depth, or holographic data representations. Different capturing systems which may be considered include multi-camera systems, stereo cameras, lidar (depth) systems, or holographic cameras. Different advanced signal processing algorithms may be involved on the sender

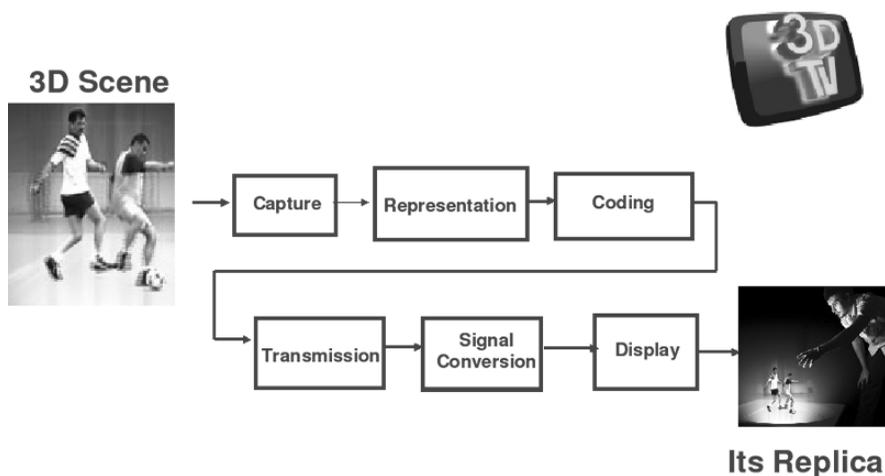


Fig. 1.2. Functional blocks of an end-to-end 3DTV system (from L. Onural, H. M. Ozaktas, E. Stoykova, A. Gotchev, and J. Watson, An overview of the holographic display related tasks within the European 3DTV project, in *Photon Management II: SPIE Proceedings 6187*, 2006)

side, including three-dimensional geometry reconstruction, depth estimation, or segmentation, in order to transform the captured data into the selected three-dimensional scene representation.

Specific compression algorithms need to be applied for the different data types. Transmission over different channels requires different strategies. The vast amount of data and user interaction for FVV functionality essential to many systems complicates this task even further. On the receiver side, the data needs to be decoded, rendered, and displayed. In many cases this may require specific signal conversion and display adaptation operations. Interactivity needs to be taken care of. Finally, the images need to be displayed. Autostereoscopic displays have already been mentioned, but there are also other more ambitious types of displays. Such displays include volumetric displays, immersive displays and, of course, holographic displays. For those who have set their eyes on the ambitious applications of three-dimensional imaging, the fully-interactive, full parallax, high-resolution holographic display is the ultimate goal. Whether or not this is achievable depends very much on the ability to efficiently handle the vast amounts of raw data required by a full holographic display and the ability to exploit the rapid developments in optical technologies.

Current end-to-end 3DTV systems require tightly coupled functional units: the display and the capture unit must be designed together, and therefore, compression algorithms are also quite specific to the system. However, it is quite likely that in future 3DTV systems, the techniques for 3D capture and 3D display will be totally decoupled from each other. It is currently envisioned that the information provided by the capture device will provide the basis for the computerized synthesis of the 3D scene. This synthesis operation will heavily utilize 3D computer graphics techniques (which are commonly used in computer animations) to assemble 3D scene information from multiple-camera views or other sets of complementary data. However, instead of synthetic data, the 3D scene information will be created from a real-life scene. Many techniques have been developed for the capture of 3D scene information. A common technique is based on shooting the scene simultaneously from different angles using multiple conventional 2D cameras. Camera arrays with up to 128 cameras have been discussed in the literature. However, acceptable quality 3D scene information can be captured by using a much smaller number of cameras, especially if the scene is not too complex.

The synthesized 3D video, created from the data provided by the capture unit, can then be either transmitted or stored. An important observation is that, 3D scenes actually carry much less information than one may initially think. The difference between 2D images and 3D images is not so much like the difference between a 2D array and a 3D array of numbers, since most objects are opaque and in any event, our retinas are two-dimensional detectors. The difference is essentially the additional information associated with depth and parallax. Therefore, 3D video is highly compressible. Special purpose compression techniques have already been reported in the literature and research

in this area is ongoing. Transmission of such data is not too different than transmission of conventional video. For example, video streaming techniques which are commonly used over the Internet can easily be adapted to the 3D case. Nevertheless, such adaptation does require some care as the usability of incomplete 3D video data is totally different than the usability of incomplete 2D video, and packet losses are common in video streaming.

In order for the display to show the 3D video, the received data in abstract form must first be translated into driving signals for the specific 3D display device to be used. In some cases, this can be a challenging problem requiring considerable processing. Development of signal processing techniques and algorithms for this purpose is therefore crucial for successful realization of 3DTV.

Decoupling of image acquisition and display is advantageous in that it can provide complete interoperability by enabling the display of the content on totally different display devices with different technologies and capabilities. For instance, it may be possible to feed the same video stream to a high-end holographic display device, a low-end stereoscopic 3D monitor, or even a regular 2D monitor. Each display device will receive the same content, but will have a different signal processing interface for the necessary data conversion.

In the near future, it is likely that multiview video will be the common mode of 3DTV delivery. In multiview video, a large amount of 2D video data, captured in parallel from an array of cameras shooting the same scene from different angles, will be directly coded by exploiting the redundancy of data, and then streamed to the receiver. The display at the receiving end, at least in the short term, will then create the 3D scene autostereoscopically. (In the long term, the autostereoscopic display may be replaced with volumetric or holographic displays.) Standardization activities for such a 3DTV scheme are well underway under the International Organization for Standardization Moving Picture Experts Group (ISO MPEG) and International Telecommunication Union (ITU) umbrellas.

Countless applications of 3D video and 3DTV have been proposed. In addition to household consumer video and TV, there are many other consumer applications in areas such as computer games and other forms of entertainment, and video conferencing. Non-consumer applications include virtual reality applications, scientific research and education, industrial design and monitoring, medicine, art, and transportation. In medicine, 3DTV images may aid diagnosis as well as surgery. In industry, they may aid design and prototyping of machines or products involving moving parts. In education and science, they may allow unmatched visualization capability.

Advances in this area will also be closely related to advances in the area of interactive multimedia technologies in general. While interactivity is a different concept from three-dimensionality, since both are strong trends, it is likely they will overlap and it will not be surprising if the first 3DTV products also feature a measure of interactivity. Indeed, since interactivity may also involve immersion into the scene and three-dimensionality is an important aspect of

the perception of being immersed in a scene, the connections between the two trends may be greater than might be thought at first.

Although the goals are clear, there is still a long way to go before we have widespread commercial high-quality 3D products. A diversity of technologies are necessary to make 3DTV a reality. Successful realization of such products will require significant interdisciplinary work. The scope of this book reflects this diversity. To better understand where each chapter fits in, it is helpful to again refer to the block diagram in Fig. 1.2.

Chapter 2 presents a novel operational end-to-end prototype 3DTV system with all its functional blocks. The system is designed to operate over a terrestrial Digital Media Broadcast (T-DMB) infrastructure for delivery to mobile receivers.

Chapters 3, 4, and 5 deal with different problems and approaches associated with the capture of 3D information. In Chap. 3, a novel 3D human motion capture system, using simultaneous multiple video recordings, is presented after an overview of various human motion capture systems. Chapter 4 shows that it is possible to construct 3D objects from stereo data by utilizing the texture information. A totally different 3D shape capture technique, based on pattern projection, is presented in detail in Chap. 5.

Representation of dynamic 3D scenes is essential especially when the capture and display units are decoupled. In decoupled operation, the data captured by the input unit is not directly forwarded to the display; instead, an intermediate 3D representation is constructed from the data. Then, this representation is used for display-specific rendering at the receiving end. Chapters 6 and 7 present examples of representation techniques within an end-to-end 3DTV system. In Chap. 6, a detailed overview of modeling, animation, and rendering techniques for 3D are given. Chapter 7, on the other hand, details a representation for the more specific case where the object is a moving human figure.

Novel coding or compression techniques for 3DTV are presented in Chaps. 8 and 9. Chapter 8 deals specifically with the compression of 3D dynamic wire-mesh models. Compression of multi-view video data is the focus of Chap. 9, which provides the details of an algorithm which is closely related to ongoing standardization activities.

Transport (transmission) of 3DTV data requires specific techniques which are distinct from its 2D counterpart. Issues related to streaming 3D video are discussed in Chap. 10. Chapter 11 discusses the adaptation of the multiple description coding technique to 3DTV.

Watermarking of conventional images and video has been widely discussed in the literature. However, the nature of 3D video data requires novel watermarking techniques specifically designed for such data. Chapter 12 discusses 3D watermarking techniques and proposes novel approaches for this purpose.

Different display technologies for 3DTV are presented in Chaps. 13, 14, and 15. Chapter 13 gives a broad overview of the history of domestic 3DTV displays together with contemporary solutions. Chapter 14 describes an

immaterial pseudo-3D display with 3D interaction, based on the unique commercial 2D floating-in-the-air fog-based display. Chapter 15 gives an overview and the state-of-the-art of spatial light modulator based holographic 3D displays. Chapter 16 discusses in detail the physical and chemical properties of novel materials for dynamic holographic recording and 3D display.

Finally, the last chapter discusses consumer, social, and gender issues associated with 3DTV. We believe that early discussion and investigation of these issues are important for many reasons. Discussion of consumer issues will help evaluation of the technologies and potential products and guide developers, producers, sellers, and consumers. Discussion of social and gender issues may help shape public decision making and allow informed consumer choices. We believe that it is both an ethical and a social responsibility for scientists and engineers involved in the development of a technology to be aware of and contribute to awareness regarding such issues.

We believe that this collection of chapters provides a good coverage of the diversity of topics that collectively underly the modern approach to 3DTV. Though it is not possible to cover all relevant issues in a single book, we believe this collection provides a balanced exposure for those who want to understand the basic building blocks of 3DTV systems from a broad perspective. Readers wishing to further explore the areas of 3D video and television may also wish to consult four recent collections of research results [1, 2, 3, 4] as well as a series of elementary tutorials [5].

Parts of this chapter appeared in or were adapted from [6] and [7]. This work is supported by the EC within FP6 under Grant 511568 with the acronym 3DTV.

References

1. *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication*. O. Schreer, P. Kauff, and T. Sikora, editors. Wiley, 2005.
2. *Three-Dimensional Television, Video, and Display Technologies*. B. Javidi and F. Okano, editors. Springer, 2002.
3. Special issue on three-dimensional video and television. M. R. Civanlar, J. Ostermann, H. M. Ozaktas, A. Smolic, and J. Watson, editors. *Signal Processing: Image Communication*, Vol. 22, issue 2, pp. 103–234, February 2007.
4. Special issue on 3-D technologies for imaging and display. B. Javidi and F. Okano, editors. *Proceedings of the IEEE*, Vol. 94, issue 3, pp. 487–663, March 2006.
5. K. Iizuka. Welcome to the wonderful world of 3D (4 parts). *Optics and Photonics News*, Vol. 17, no. 7, p. 42, 2006; Vol. 17, no. 10, p. 40, 2006; Vol. 18, no. 2, p. 24, 2007; Vol. 18, no. 4, p. 28, 2007.
6. L. Onural. Television in 3-D: What are the prospects? *Proceedings of the IEEE*, Vol. 95, pp. 1143–1145, 2007.
7. M. R. Civanlar, J. Ostermann, H. M. Ozaktas, A. Smolic, and J. Watson. Special issue on three-dimensional video and television (guest editorial). *Signal Processing: Image Communication*, Vol. 22, pp. 103–107, 2007.