AN ASSESSMENT OF 3DTV TECHNOLOGIES

Levent ONURAL
Dept. of Electrical and Electronics
Bilkent University
Ankara, Turkey
onural@bilkent.edu.tr

Thomas SIKORA
Institut für Telekommunikationssysteme
Technische Universität Berlin
Berlin, Germany

Jörn OSTERMANN
Institut für Informatiosnverarbeitung
University of Hannover
Hannover, Germany

Aljoscha SMOLIC
Fraunhofer-Inst. für Nachrichtentechnik
Heinrich-Hertz-Institut
Berlin, Germany

M. Reha CİVANLAR
Dept. of Computer Eng
Koç University
Istanbul, Turkey

John WATSON
Dept. of Engineering
University of Aberdeen, King's College
Aberdeen, Scotland

ABSTRACT

Stereoscopic 3D viewing techniques are almost as old as their 2D counterparts: experimental stereoscopic 3DTV immediately followed the invention of TV. Holography is a newer technology compared to stereoscopy, and there are indicators that satisfactory holographic 3DTV may be feasible. Another candidate technology for 3DTV is integral imaging. Holography and integral imaging provide true full parallax 3D displays in the ideal case. All these technologies have their own distinct features, advantages and problems. Interest in all forms of 3DTV has been significantly increasing both in research and commercial communities. An integrated 3DTV system naturally has different components: capturing of 3D moving scenes, their representation, compression and transport, and finally display are the main building blocks. Naturally, the consumer attitude and the related social issues will be rather centred around the display and interaction. 3D scenes can be captured by various means, for example, by using many cameras simultaneously. Furthermore, it is desirable to serve all types of 3D displays with different capabilities. Therefore, It is envisaged that scene capture and display operation will be decoupled in future 3DTV systems: captured scene information will be converted to abstract representations (and maybe stored) using computer graphics techniques, and the display (and observer) will interact with this intermediate data. It is natural to extend conventional video compression techniques to 3D video signals by exploiting the inherent redundancies. Coding of 3D video signals is attracting research interest and related standardization activities are ongoing in bodies like ISO-MPEG. Digital transmission, using adapted streaming techniques is another research area. Autostereoscopic, holographic and volumetric displays have been demonstrated and used. Signal processing techniques are employed to find the technology-dependent display driver signals to get the 3D images from abstract 3D scenes.

INTRODUCTION AND HISTORICAL OVERVIEW

The ultimate goal of the viewing experience is to create the illusion of a real environment in its absence. If this goal is fully achieved, there is no way for an observer to distinguish whether or not what he sees is real or an optical illusion.

Due to its ease and technological feasibility in earlier times, 2D representations of images have been with us since the beginning of history in the form of paintings and drawings. Capturing the sense of depth in simple 2D drawings has been a challenge; art historians are well aware of developments in art which led to drawing techniques using perspective techniques. Photography was publicly introduced in 1839 by Sir John Herschel; however, the underlying optical process was known about three centuries before that date [1]. Since then, 2D still imaging has been improved to near its limits giving us the beautiful pictures we see around us. Animation of 2D pictures to achieve movies was documented in 1867 in a US patent about a device called ``zoopraxiscope" invented by William Lincoln [2]. Observing images from remote places was accomplished by the invention of television as early as 1920s (Edouard Belin and John Logie Baird) [3]. The high quality viewing experience in today's digital TV sets and movie houses is a natural consequence of continuing scientific and technological developments, improvements and inventions in this field. Of course, the driving force behind all this development is the never ending consumer demand for a better viewing experience, the curiosity and the talents of those who provide the technological basis, and the entrepreneurial skills and attempts to satisfy such demands.

As soon as the photography and its motion picture equivalent were invented, stereoscopic 3D immediately followed. Indeed, it is known that a mirror device was used in 1838 to deliver stereoscopic 3D images by Sir Charles Wheatstone [4]. By 1844, stereoscopic viewing was popular both in Europe and in USA. Similarly, the stereoscopic 3D counterparts of motion picture and television quickly became a reality after the invention of their 2D counterparts: the concept of stereoscopic cinema appeared in the early 1900's, and stereoscopic TV was proposed in the 1920's. By 1950, 3-D movies became quite popular. 3D movie theaters spread all over the world with their high resolution format, and gave audiences a highly satisfactory stereoscopic 3D experience. Although experimental 3DTV broadcast dates back as early as 1953, the first commercial 3DTV broadcast took place in 1980 in the USA [5].

An overview of 3D exhibitions in different parts of the world between 1985-1996 and technologies presented in those exhibitions are given in [8].

Stereoscopy is rather simpler; its fundamental operational principle is based on the human visual system and perception. However, most stereoscopic systems create mismatches between various 3D cues in human perception, and thus create discomfort while viewing. Indeed, most of the current research in stereoscopic 3D is targeted to overcome such problems [4,5].

Even though stereoscopy has been known for a long time, there are other 3D imaging techniques, and some of them are also known for a long time. Based on scientific developments in optics and diffraction theory dating back to the early 1600's, the principles of holography were established in 1948. [6]. The first off-axis holograms were created in the early 1960s when lasers became available. Digital holography techniques followed, and eventually holographic cameras appeared. Experimental holographic motion pictures appeared for the first time in 1989 [6]. Recent developments in this field strongly hint at successful holographic 3DTV displays being produced in the near-future.

Another technology for 3D imaging is usually referred as ``integral imaging'' and known since 1908 [6,7]. The basics of integral imaging can be explained as capturing many 2D pictures of an object simultaneously from different angles, and then optically projecting the pictures back to the geometric location of the object, in its absence, to create the 3D image. Lenslet arrays (micro-lens arrays) are generally used in both capture and reconstruction. Extension of the technique for motion pictures and TV is possible. Integral imaging is a strong candidate for next generation of 3DTV.

Holography and integral imaging provide true full parallax 3D displays. Unlike stereoscopy, their principles are not based primarily on human visual perception, but on the principle of duplicating the physical light distribution in the viewing space in the absence of the original objects. The quality of the generated 3D image is, therefore, based on the success in duplicating the physical properties of the original light. Scientific and technological developments in both fields have been significant and the quality of displays has been improving. Similarly, the problems in stereoscopy are being solved with the advances in autostereoscopic multi-user systems. The ultimate goal is to provide the viewer with the freedom to move and change his or her viewing direction while interacting with the 3D image and virtual environment, together with a perception of the vivid colors and sharpness that we experience in real life. The association of still 3D imagery with 3D motion pictures and 3DTV are similar to the 2D case: if the visual information can be updated fast enough, motion will be observed, and if that data can be electronically transmitted, we get 3DTV. The difference is in the detail in the technology used to capture, represent, transmit, and display such pictures.

As seen from the brief historical overview above, the 3D imaging technologies have been known and utilized for a long time; indeed, it will not be grossly wrong to state that the 2D and 3D technologies have been developed in parallel. Yet, it is a simple observation that the popularity of 2D products in any form surpasses their 3D counterparts by far. The reasons for this imbalance, and the basic underlying consumer attitude and preferences should be well understood to overcome this unfavorable situation for 3D. The history is full of unsuccessful entrepreneurial attempts in the form of business failures in 3D imaging. As the reasons of such failures are understood by analyzing the consumer behavior, and as the technology provides the solutions to the problem areas, there is no doubt that the 3D viewing will be the choice of the future. Such a future will provide a completely new experience. Any associated social and psychological impact remains unknown at this time.

However, it is certain that the recent interest in 3D imaging, both in society, and in the research community is increasing significantly. An indicator is the volume of scientific papers, news articles, and patents in these fields.

A collection of historical pictures in 3D stereoscopic imaging (both still and motion) is presented in [14].

A stereoscopic 3D-HDTV System was reported in 1999 in NHK-STRL annual report [9,16]. The report mentions the regular problems associated with stereoscopy, and describes subjective evaluation tests targeted at overcoming such problems. It is claimed that two factors, "sensation of reality" and "ease of viewing" are extracted from such tests. With improvements in these factors, this study concluded that 3D images were better than 2D in terms of sensation-of-reality, but scores for ease-of-viewing varied depending on the image content. A discussion of future 3DTV systems is presented including autostereoscopic, holographic and integral-imaging-based systems. Capturing techniques for 3D scenery is also covered, and a description of a 3D camera (1998), based on infrared sensors to detect depth, is presented. Some test results on visual and psychological effects associated with wide-screen display systems are also given.

Another Korean broadcasting experiment in 3D-HDTV was the broadcast of 2002 FIFA World Cup within activities in 3D-HDTV project [13]. The project spanned human visual fatigue studies, stereoscopic cameras, video multiplexer-demultiplexer, receiver, coding, related image processing techniques based on MPEG-2 and MPEG-4, etc. Different stereoscopic cameras were tested. The activities involved 10 demo rooms with 50 seats and a 300 in. screen; it is claimed that the demo rooms were visited by about 571,000 visitors during the events. The stereoscopic viewing was via polarizing glasses.

Perceptual evaluation of 3DTV displays and system requirements based on such evaluations are presented in [10]. The focus is only on stereoscopic displays, and autostereoscopic systems with multiple viewers. Rather immature holographic or integral-imaging-based displays are omitted. On the capture side, dual camera (stereoscopic), single camera assisted with a depth camera, and a single camera with 2D-to-3D conversion are considered. Captured data is compressed and delivered to the displays. Evaluation paradigms are discussed, and in particular applicability of 2D video assessment techniques to 3D are questioned. It is concluded that 3D experience is quite different, and therefore, must be assessed based on criteria that fit better to 3D. Six major viewing artifacts for the stereoscopic case are listed and discussed.

ATTEST was a project on 3DTV funded between 2002-2004 by the EC. A full 3DTV processing chain has been realized and demonstrated in the European ATTEST project [17]. The result is a backward compatible (to classical DVB) approach for 3DTV. In this context compression of depth data has also been investigated. It has been found that depth data can be very efficiently compressed using standard video codecs such as H.264/AVC [18]. From standards point of view the realization of the ATTEST concept for 3DTV only requires minor additions on the Systems level of MPEG-4. These are currently under investigation and may provide an interoperable solution for 3DTV broadcast in the very near future. This concept for depth based 3D rendering is easily extended to N views, shown in [19]. Depending to the user position a simple switching to the nearest original view with depth (or pair of views with disparity/depth) is possible. This extends the navigation range in front of the screen with the number of cameras used. For some application scenarios such as 3DTV broadcast this implies compression and transmission of multi-view video, which is an ongoing work item in MPEG standardization activities.

Newer generation of 3DTV techniques are targeted to decouple the image capture and image display components further: in such systems, the captured scene, by some means, is first converted to an abstract 3D moving scene using such aids like wire-mesh models and other representation techniques. The 3D scene is then rendered at the display side depending on the display technique employed. Based on human perception and physical properties and the technology of the display, there are many different ways of rendering the captured 3D info. One such system, based on scanning different depth slices of a 3D scene by holographic means of reproducing each slice in a time sequential fashion is presented in [12].

A PC-based stereoscopic interactive video system to give the sensation of walking through a pre-recorded 3D environment is presented in [11].

A 3D videoconference application is described in a patent document [15]. The 3D image is captured by an array of video cameras. Digitized video data is computer processed and the resultant data is transmitted. 3D data collected from all such tele-conference attendees are collected to form a single 3D image which is then transmitted to all locations. Received 3D video data is displayed using a 3D projection system.

A paper published in 1995 [20] describes the state of 3DTV research at that time. An overview of human factors is presented; stereoscopic 3DTV systems related issues are discussed; bandwidth and its possible reduction through coding are included.

After a brief general introduction about early anaglyphic broadcasts in Europe in the early 1980s, more advanced two-channel PAL demonstrations both from Europe and Japan in 1983, 1985 and 1987 are mentioned in [21]. Then a technological overview of research in Europe is presented, including psycho-optic aspects and signal processing issues. An overview of European COST 230 ``Stereoscopic Television" is also given, together with the RACE DISTIMA project. Developments in Japan and USA are also briefly presented.

An end-to-end distributed scalable 3DTV system, consisting of an array of cameras, clusters of network connected PCs, and a multi-projector display is developed and implemented by Mitsubishi Electric Research Laboratories (MERL) [22]. Multiple video streams are individually encoded and transmitted over broadband networks. The display is based on lenticular technology. Design choices and tradeoffs are presented.

SCENE CAPTURE AND REPRESENTATION FOR 3DTV

Three-dimensional television starts with acquiring the dynamic, real-world scene in some suitable digital representation. In contrast to conventional TV, however, not only the visual appearance of the scene must be recorded, but 3DTV requires to additionally acquiring also complete shape information in order to enable looking at the scene from different view points. The scientific challenges are twofold: 3D geometry of scenes in motion must be acquired, while the original visual appearance of the scene may not be altered.

A number of different technologies have the potential to meet these requirements. On one extreme we find purely image-based approaches using several conventional cameras. Computer vision and computer graphics techniques are then used to describe the recorded scene such that a user can look at it from different angles. On the other extreme we find active holographic techniques. Recent advances in CCD and CMOS imaging technologies show promise to enable direct digital hologram acquisition in the future. Here we provide an overview of the image-based approaches.

The conceptually simplest solution to scene capture is to place a camera at each location from which the scene should be looked at and to display the appropriate two views to the human observer [23]. However, this might require an infinite amount of cameras. Typically a set of 2 to 20 cameras are used in a multi-camera recording system which is a calibrated recording setup consisting of cameras delivering synchronized video streams.

For calibration, a point light source is moved in the entire space that all the cameras look at. Calibration information like the internal and external parameters (position, orientation, lens information) of the cameras is computed using the recorded videos [24]. Algorithms for automatically calibrating cameras as they record an arbitrary scene are still an active research area [25].

Due to calibration, the location in the camera image of a 3D point of the scene can be computed for all camera images. Using the image coordinates of a 3D point in at least two camera images, the inverse problem can be solved: What are the 3D coordinates of this point? In a first step, feature points like corners are located in a first image [26]. In a second step the location of each feature point using the texture of the feature point in the first images is located in the other images [27,28]. This search is simplified by the calibration information which defines for each of the other images just one line in each image where the point has to be located [29]. These feature points may also be tracked over time in order to increase the reliability of the estimated 3D coordinates [30,31].

As soon as the 3D coordinates of the points of the scene are identified, a 3D surface model of the scene is created. The surface of an object is described using a mesh of polygons where the vertices of the mesh are located at the estimated 3D coordinates. Important alternative representations are triangle meshes, NURBS [32] and subdivision surfaces [33]. Subdivision surfaces offer a good compromise between an inherently non-smooth polygonal mesh representation and NURBS surfaces which are limited by topological restrictions [34]. Subdivision surfaces allow representation of arbitrary topology and any fine detail with a controllable smoothness.

In a final step, the image is projected onto the 3D model defining for each surface patch the look or texture. In advanced systems, the texture of several or all images where the patch can be seen is attached. Hence, each patch has several texture maps enabling a more realistic rendering of the object for different viewpoints. As the number of available images increases, the 3D geometry can be of less precision. There are several approaches of representing an object starting from precise 3D shapes with just one texture up to many images of the object without explicit 3D shape [23,35,36]. A 3D model can be rendered from an arbitrary viewpoint using well-known rendering algorithms based on OpenGl, Direct3D or other graphics libraries.

CODING AND STANDARDIZATION ACTIVITIES

As shown in the previous section there are different data types are used for the different 3D scene representations in the context of 3DTV. Having defined the data, efficient compression and coding is the next block in the 3D video processing chain, and that is the scope of this section. There are many different data compression techniques corresponding to different data representations. For example, there are different techniques for 3D meshes, depth data, multiple view video, etc. However, the level of maturity varies largely. There is a strong relation to the age, level of maturity and the (commercial) usage of the corresponding data representation.

One class of data is related to compression of any kind of pixel data, such as video, stereo video, multi-view video, but also associated per-pixel depth data, etc. This wide field is partially well established but partially also very innovative, and in any case highly current and newsworthy.

Compression of classical 2D video for instance has been studied very intensely over decades by a very large number of researchers and institutions. As result the latest generation video codecs such as standard H.264/AVC provide excellent performance. Scalability features will be added to H.264/AVC in the current SVC activity in MPEG. Nevertheless, there is still room for improvement of basic 2D video coding. These include a better pre-analysis and exploitation of semantics, as well as wavelet approaches.

Similar conclusions can be drawn for stereo video, which can be regarded as first order extension. Commercial usage is not that large as for 2D video but the technology is quite mature. However, segmentation and object-based representation play a more important role for stereo video, and these fields still represent major algorithmic challenges.

The N-dimensional extension called multi-view coding (MVC) is relatively young; however, it currently receives very large attention. MPEG issued a related "Call for Proposals" that was evaluated in January 2005. This will lead to a new dedicated standard for MVC. MVC is a basic component for certain 3DTV and free viewpoint video systems.

The nature of depth and disparity data is similar to 2D video (i.e. temporal succession of matrices of integers). Compression of such data has also been studied to some extend. Available standards as MPEG-4 already allow for compression and transmission of such data. However, also in this area there is still room for improvement, using specific algorithms that better exploit the (e.g. statistical) nature of depth data. The concept of layered depth images (LDI) can be regarded as a natural extension to N views with depth of the same scene. This type of data representation is relatively young but highly interesting for certain 3DTV applications. There is also a strong relation to MVC. Principles from depth compression can be extended to LDI, but further improvement can be expected exploiting e.g. inter-view dependencies as done in the case of MVC.

A light-field representation is also a relatively young data type which stores the images of a scene from different angles. So far mainly static light-fields have been investigated. Dedicated compression algorithms have been presented in some pioneering work. In principle there is a strong relation to MVC. For instance dynamic light-field compression is handled in MPEG as a specific case of MVC. The practical relevance of very dense dynamic light-fields is still questionable. Nevertheless, significant improvements of compression performance using dedicated algorithms can be expected.

3D meshes are widely used in computer graphics. Compression of such data has therefore also been widely studied. However, further improvements are possible especially for progressive and dynamic (i.e. time varying) meshes. For the latter, there is a related activity in the SNHC group of MPEG. Dynamic meshes have not received much interest in the past. Significant improvements can be expected by incorporating basic principles from video coding.

A point cloud representation is an alternative to classical 3D meshes. Such a representation might be very interesting for certain 3D video applications. Pioneering work on compression and streaming has been presented, but there seems to be a lot of room for improvement.

Holographic signals have so far not been used in multimedia applications, although highly interesting for 3D displays. The commercial relevance of such a data representation is still uncertain. Naturally, compression is not yet studied in detail. This is an open research field where a lot of work would have to be done if such data become relevant.

Multiple description coding and channel adaptation also currently receives significant attention. Here it is shown that improvements are possible for specific application fields if some of the basic coding paradigms of available standard video coding are abandoned. This research direction should be further pursued with specific focus on 3D video data.

As for any type of media, security and rights management is also an important issue for 3D video. Some research has been done for classical 3D models. However, there still needs to be done a lot and for other data this is still an open field.

In general conclusion we may state that the very diverse research area of 3D video compression is highly active and relevant at the moment. Market relevance and interest of manufacturers, content providers and users in 3D video systems are growing rapidly. However, there are still important challenges that need to be resolved. One of the goals of the European Community funded 3DTV project is to integrate the European research efforts in 3D video compression to ensure a strong European participation in this highly relevant future market.

TRANSPORTING 3D VIDEO

Determination of the best techniques for transporting 3DTV data over communication networks in real-time requires a thorough investigation of several classical communication techniques together with their adaptation to the unique requirements of this new application. Experiences gained in the early implementations of 3DTV systems, as discussed in the previous sections, are extremely important in reaching a clear understanding of 3DTV transport issues, and therefore must be carefully studied.

It is logical to expect that the transport infrastructure for any new communication application will be based on packet network technology and employ the Internet Protocol (IP) suites. The IP architecture is proving to be flexible and successful in accommodating a wide array of communication applications as can be seen from the ongoing replacement of classical telephone services by voice over IP applications. Transport of the TV signals over IP packet networks seems to be a natural extension of such applications. Video-on demand services, both for news and for entertainment applications, are already being offered over the Internet. Also, 2.5G and 3G mobile network operators started to use IP successfully to offer wireless video services. Therefore, we visualize a 3DTV transport system based on packet network technology and IP. Systems for streaming 3D video over the Internet can be built based on the vast experiences obtained in 2D applications. However, 3D video can have a much larger bandwidth demand and very specific dependency structures in the transmitted data. The 3DTV modalities used for 3DTV have significant effects on streaming system implementations. These modalities, particularly viewed from the transmission aspect, may be summarized using a linear spectrum. At the leftmost side of this spectrum are the techniques for completely synthetic video generation, that is, techniques based on computer graphics. As we move to the right hand side of the spectrum, we can see the techniques that mix graphics with real images, such as those that use depth information together with image data for 3D scene generation [47]. Purely image based rendering techniques [48], light-fields [49], are located at the right side of this spectrum. And, at the rightmost edge, we can put holographic video. Clearly, as we move along this spectrum of modalities, the transmission issues vary a great deal. For example, graphic techniques, do not require a very large transmission bandwidth, but their loss tolerance may be extremely low. Several techniques for loss resilient transport of synthetic video can be found in the literature, e.g. [50,51]. Purely image based techniques are much more loss tolerant, but their bandwidth demand is much larger. At the rightmost end of the spectrum, the bandwidth requirements may exceed anything that is available at the current state of the data transmission technology.

The large bandwidth demand of the image based and holographic techniques makes the use of efficient compression a vital necessity. As discussed in the previous sections, several effective compression techniques for multi-view video have been developed and this continues to be an active research area. From the transmission viewpoint, two important aspects of the use of compression are the reduced loss resiliency and data dependency. As the redundancy in the data is removed, so does the inherent loss resilience. And, significant compression gain in multi-view video compression is obtained through interview prediction, but this creates a dependency between the views. Nevertheless, the techniques for handling 2D compressed video transport over lossy networks are well developed and similar approaches are applicable to 3DTV transport. These include use of application layer framing and layered coding with unequal error protection. Techniques for the concealment of packet loss effects become very important as well. Loss concealment in 3D can't be accomplished by a straightforward extensions of the techniques used for 2D video. New approaches have been one of the active research areas [52].

Another aspect of 3DTV video which does not exist in its 2D counterpart is the dependency of the displayed video to the viewpoint of the viewer. The video must be adjusted when the viewer moves around, changing his or her viewpoint of the display. Otherwise, the displayed scene will be quite unrealistic. Particularly for image based techniques; however, this requires transmission of a multitude of views to the end points, multiplying the bandwidth requirements by many factors. Efficient networking techniques for multi-view video delivery over multicast networks is therefore an active research area [53].

Finally, cross layer approaches, where several layers of the communication architecture, from application to physical, are considered together, and jointly optimized, have recently shown to be very successful in 2D applications. Their extension to 3D looks very promising. This approach is particularly important in wireless applications, which may be one of the leading applications of 3DTV, because of the tendency of the wireless operators to feature new applications much earlier than their wired counterparts.

3DTV DISPLAY TECHNOLOGIES

The display is the last, but definitely not least, significant aspect in the development of 3D vision. As has already been outlined, there is a long chain of activity from image acquisition, compression, transmission and reconstruction of 3D images before we get to the display itself. However, the display is the most visible aspect of the 3DTV and is probably the one by which the general public will judge its success. The concept of a three-dimensional display has a long and varied history stretching back to the 3D stereo-photographs made in the late 19th century through 3D movies in the 1950's, holography in the 1960's and 70's and 3D computer graphics and virtual reality of today.

The need for 3D displays and vision grows in importance by the day, as does the number of applications such as scientific visualization and measurement, medical imaging, telepresence, gaming, as well as movies and television itself. Many different methods of 3D displays have been presented over the last few decades [8], but none has been able to capture the mass market. Much of development in 3D imaging and displays of the latter end of the 20th century was spurred on by the invention of holography, and this was certainly the catalyst which led to some of the significant advances in autostereoscopic and volumetric methods, whereas, advances in techniques of virtual reality have helped to drive the computer and optics industries to produce better head-mounted displays and other 3D displays.

The main requirement of a 3D displays is to create the illusion of depth or distance by using a series of depth cues such as disparity, motion parallax, and ocular accommodation [5,10]. Additional cues are also needed for image recognition. Conflicting cues are one of the leading causes for discomfort and fatigue when viewing 3D displays. The form that such displays would take is one aspect which needs considerable thought and is a major concern in consumer acceptance. Will the consumer want to see the "Star Wars" image projection out of a central table or will a flat panel in the corner of a room be the norm? It may well be that the application will drive the technology. Important aspects to be considered include image resolution, field of view, brightness, whether they are single or multi-user, viewing distance and cost.

The technologies being pursued for 3D display can be broadly divided into the following categories, as shown in Figure 1 (although there are various other methods of classification used and the terminology is not always clear):

- Holographic displays [e.g. 54-56]
- Volumetric displays [e.g. 57,58]
- Autostereoscopic displays [e.g. 59-62]
- Head mounted displays (HMD) [e.g. 63]
- Stereoscopic displays

The term "autostereoscopic", strictly speaking, describes all those displays which create a stereoscopic image without any special glasses or other user-mounted devices and in this respect might be considered to include holographic, volumetric and multiple image. However, we restrict the use of the term to cover displays such as binocular, multi-view and holoform systems where only multiple two-dimensional images across the field of view are considered. Autostereo-systems are limited by the number of viewers and eye or head tracking is usually needed. In holographic displays the image is formed by wave-front reconstruction, and includes both real and virtual image reconstruction. Holography is at present handicapped by the vast amount of information which has to be recorded, stored, transmitted and displayed, putting severe constraints on the display technology employed. Furthermore, holography can be deployed in reduced parallax (e.g. stereoholography or lenticular) systems, which relax some of the constraints. Volumetric displays form the image by projection within a volume of space without the use of light interference, but have limited resolution. Head mounted displays such as those using, for example, liquid-crystal-on-silicon (LCOS) devices or retinal scanning devices (RSD) are unlikely to fine mass-market acceptance because of the user discomfort similar to motion sickness and the public reluctance to wear devices, but may find some well-defined niche markets. The more conventional stereo-technologies, all require the use of viewing aids such as red/green or polarizing glasses.

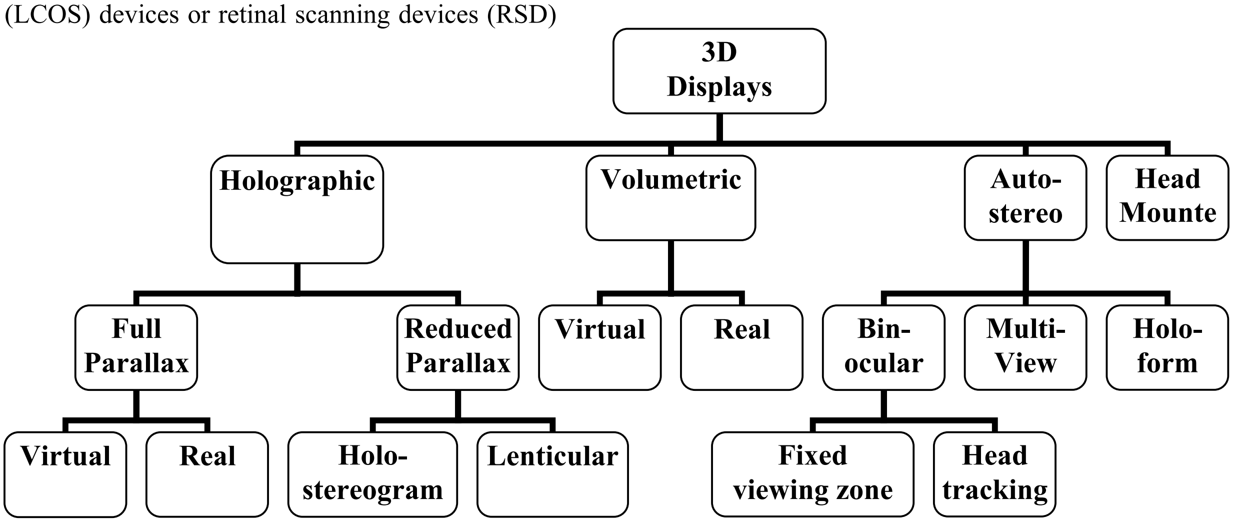

Figure 1: Classification of 3D Display Techniques.

Clearly no display method is without its problems or limitations. The development paths which have to be followed before a full 3D display can be realised are very complex. Given the current state-of-the-art, non-holographic displays, such as volumetric or autostereo, are in a more advanced state of development and it is felt that they are more likely to reach the market place in a shorter time frame. A full, large area, interactive, colour holographic display, which is thought by many to be the ideal goal, requires the parallel development of many essential areas of technology before it can be brought to fruition.

As an example of the development path which may take place and steps that need to be taken on the way, we can envisage the development of a large, wide-angle, full colour, full parallax, moving, interactive holographic display for television. We can draw a rough road map through to completion of such an objective. It is clear that to reach such a goal a series of incremental improvements are needed on the way. To reach this stage, though, requires significant progress to be made in the development of support technologies. For example, a large display of say 100 mm diagonal will need dramatic improvements in VLSI techniques to enable an SLM to be manufactured with sufficient pixel resolution. If the oft-quoted "Moore's Law" continues to apply then it could still be more than 8 years before a display of less than micron pixel size is achieved. An array of SLM's requires advances in interconnection technology and software required to drive them. Colour displays require development of compact, safe lasers or LED's with sufficient coherence and power. Another important issue is that of parallax. It has often been said that someone viewing a hologram for the first time only notices the presence of vertical parallax when "jumping up and down with excitement"! It is true that in the case of 3DTV or movies, the viewer will normally be seated and unaware of vertical parallax. It is probable that such systems could usefully sacrifice vertical parallax. However, in an operating theatre the argument for loss of parallax is not so valid. A similar timeline could be drawn for a purely autostereoscopic or volumetric display. However, it is felt that the advances currently being made in autostereo displays suggest that a multi-viewer, high resolution, bright display could be achieved two or three years earlier than a holographic one.

The pursuance of the goal of a full 3D display for TV or other vision applications is an ever-expanding field of endeavour. Many approaches have been outlined and discussed, from simple stereo with red/green glasses through to full parallax holography. What technology is applied on a given occasion will largely depend on the application. For example, it maybe that a full parallax, full colour, interactive holographic display would be used in air traffic control but that an autostereo-display is more appropriate for low level CAD applications. What is clear is that no single approach is likely to dominate and it will be the application which will determine which technology is adopted.

SIGNAL PROCESSING ASPECTS OF HOLOGRAPHIC 3DTV

Image capture and image display will most likely be decoupled in future 3DTV systems. There will be a need to convert abstract scene representations to display driver signals. For holographic displays, diffraction and propagation effects must be taken care of. Therefore, it is expected that signal processing issues will play a fundamental role in achieving 3DTV operation. Two fundamental problems are digital computation of the optical field due to a 3D object, and finding the driver signals for a given optical device so as to generate the desired optical field in space [64]. The discretization of optical signals leads to several interesting issues; for example, it is possible to violate the Nyquist rate while sampling, but still maintain full reconstruction [65]. The fractional Fourier transform is another signal processing tool which finds application in optical wave propagation [66].

EUROPEAN 3DTV NETWORK OF EXCELLENCE

A project, with acronym 3DTV has been active since September 2004. The project is funded by European Community and conducted by a consortium of 19 institutions from seven countries, coordinated by Bilkent University. There are about 200 researchers contributing. The consortium has a multidisciplinary nature, and all aspects of 3DTV outlined above, and other issues like consumer behaviour and social impact are also investigated. The consortium conducts joint research on all technical aspects of 3DTV, and targets a long-term durable integration of its researchers via various integration activities. More information of consortium activities can be found in [67,68].

CONCLUSIONS

3DTV techniques have its roots in history. Successful 3DTV systems require a delicate coupling of various technical components, and therefore, multidisciplinary in nature. It is quite possible that future 3DTV systems will have decoupled scene capture and display components, with abstract representation of 3D scenes based on computer graphics tools. Signal processing will convert basic captured 3D scene signals to appropriate signals to drive various kinds of 3DTV displays, ranging from various variants of stereoscopy to well advanced holographic ones. Current research in the field is alive and increasing its momentum.

ACKNOWLEDGEMENT

This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

REFERENCES

[1] http://www.rleggat.com/photohistory/

[2]http://inventors.about.com/library/inventors/blmotionpictures.htm

[3] http://www.tvhistory.tv/early_inventors.htm

[4] C. Fehn, 3D TV Broadcasting, Chapter 2 in O. Schreer, P. Kauff and T. Sikora (Editors), 3D Videocommunication, Wiley, 2005.

[5] State-of-the-art in Human Factors and Quality Issues of Stereoscopic Broadcast Television, W. A. IJsselstein, P. J. H. Seuntiens and L. M. J. Meesters, ATTEST Proj. Deliverable 1, 2002.

[6] Technical Committee 4, Signal Processing Issues in Diffraction and Holography: A Survey, 3DTV NoE, 2005.

[7] G. Lippmann, La Photographie Integral, Comptes Rendus, Academie des Sciences, vol 146, pp 446-452, 1908

[8] M. Starks, 3D for the 21st Century: The Tsukuba Expo and Beyond, 3DTV Corporation, 1996

[9] NHK Annual Report, 3-D Hi-Vision (HDTV) System, 1999.

[10] L. Meesters, W. A. Ijsselsteijn, P. J. H. Seuntiens, A Survey of Perceptual Evaluations and Requirements of Three-dimensional TV, IEEE Tr. On CSVT, vol 14, No 3, pp 381-391, 2004

[11] A. J. Woods, D. Offszanka, G. Martin, A PC-Based Stereoscopic Video Walkthrough, SPIE Vol 3639, pp 306-312, 1999.

[12] J-C. Grossetie, Computer-assisted method and device for restoring three-dimensional images, US Patent 6,621,605. 2003.

[13] N. Hur, C. Ahn, C. Ahn, Experimental Service of 3DTV Broadcasting Relay in Korea, SPIE vol 4864, 2002

[14] M. Starks, History of 3DTV, http://www.3dmagic.com/3dtvpicweb/index.htm

[15] D. A: Monroe, Virtual Video Conferencing System, US Pat No 6,545,700, 1997.

[16] N. Hur, G. Lee, W, You, J. Lee, C. Ahn, An HDTV-Compatible 3DTV Broadcasting System, ETRI Journal, vol 26, No 2, pp. 71-82, 2004.

[17] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. Ijsselsteijn, M. Pollefeys, L. Vangool, E. Ofek, and I. Sexton, "An Evolutionary and Optimised Approach on 3D-TV", Proc. of IBC 2002, Int. Broadcast Convention, Amsterdam, Netherlands, Sept. 2002.

[18] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audio-Visual Services", 2003.

[19] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-Quality Video View Interpolation Using a Layered Representation", SIGGRAPH04, Los Angeles, CA, USA, August 2004.

[20] T. Motoki, H. Isono and I. Yuyama, Present Status of Three-dimensional Television Research, Proceedings of IEEE, vol 83, no 7, pp 1009-1021, 1995.

[21] S.Pastoor, Human Factors of 3DTV: An Overview of Current Research at Heinrich- Hertz-Institut Berlin, IEE Colloq. On Stereoscopic Television, pp 11/1-11/4, 1992.

[22] M. Wojciech and H. Pfister, 3D TV: A Scalable System for Real-Time Acquisition, Transmission and Autostereoscopic Display of Dynamic Scenes, Mitsubishi Electric Research Laboratories Report TR2004-67, 2004.

[23] M. Levoy, P. Hanrahan, "Light Field Rendering", Proc. ACM SIGGRAPH, pp. 31-42, August 1996.

[24]Easy Calibration Toolbox http://www.cis.upenn.edu/~sequence/research/downloads/EasyCal/

[25] Faugeras Olivier, Luong Quang-Tuan, Maybank S.J., "Camera Self-Calibration: Theory and Experiments", Proceedings of the 2nd European Conference on Computer Vision, 321-334, 1992

[26] Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In 4th Alvey Vision Conference, S. 147–151. 1988.

[27] Faugeras, O., Fua, P., Hotz, B., Ma, R., Robert, L., Thonnat, M., Zhang, Z.: Quantitative and Qualitative Comparison of Some Area and Feature-based Stereo Algorithms. In International Workshop on Robust Computer Vision: Quality of Vision Algorithms, S. 1–26. 1992.

[28] L. Falkenhagen, Depth estimation from stereoscopic image pairs assuming piecewise continuos surfaces", European Workshop on combined Real and Synthetic Image Processing for Broadcast and Video Production, 1994

[29] Torr, P. H. S., Zisserman, A.: Robust Parameterization and Computation of the Trifocal Tensor. Image and Vision Computing, Bd. 15(3):591–605, 1997.

[30] Shi, J., Tomasi, C.: Good Features to Track. In IEEE Conference on Computer Vision and Pattern Recognition, S. 593–600. 1994.

[31] Zisserman, A., Fitzgibbon, A., Cross, G.: VHS to VRML: 3D Graphical Models from Video Sequences. In IEEE International Conference on Multimedia and Systems, Bd. 1, S. 51–57. 1999.

[32] Piegl L. and Tiller W., The NURBS Book, Springer 1997.

[33] Zorin D., Schröder P., and Sweldens W., "Interpolating Subdivision for Meshes with Arbitrary Topology", SIGGRAPH '96 Proc., pp. 189-192, 1996.

[34] Zorin D., Schröder P., Levin A., Kobbelt L., Sweldens W., and DeRose T., Subdivision for Modeling and Animation, SIGGRAPH 2000 Course Notes, 2000.

[35] W. C. Chen, J. Y. Bouguet, M. H. Chu, R. Grzeszczuk, "Light Field Mapping: Efficient Representation and Hardware Rendering of Surface Light Fields", Proceedings of ACM SIGGRAPH, pp. 447-456, 2002.

[36] D. Vlasic, H. Pfister, S. Molinov, R. Grzeszczuk, W. Matusik, "Opacity Light Fields: Interactive Rendering of Surface Light Fields with

View-dependent Opacity", Proc. of 2003 symposium on Interactive 3D graphics, pp. 65-74, 2003.

[37] A. Smolic, and D. McCutchen, "3DAV Exploration of Video-Based Rendering Technology in MPEG", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 14, No. 3, pp. 348-356, March 2004.

[38] ISO/IEC JTC1/SC29/WG11, "Applications and Requirements for 3DAV", Doc. N5877, Trondheim, Norway, July 2003.

[39] ISO/IEC JTC1/SC29/WG11, "Report on 3DAV Exploration", Doc. N5878, Trondheim, Norway, July 2003.

[40] ISO/IEC JTC1/SC29/WG11, "ISO/IEC 14496-16/PDAM1", Doc. N6544, Redmont, WA, USA, July 2004.

[41] K. Mueller, A. Smolić, M. Kautzner, P. Eisert, and T. Wiegand, "Predictive Compression of Dynamic 3D Meshes", Proceedings of the IEEE International Conference on Image Processing (ICIP) 2005, Genova, Italy, September 11-14, 2005.

[42] S. Würmlin, E. Lamboray, O. Staadt, and M. Gross, "3D Video Recorder: A System for Recording, Processing and Playing Three-Dimensional Video", Computer Graphics Forum 22 (2), Blackwell Publishing Ltd, Oxford, U.K., pp. 181-193, 2003.

[43] S. Würmlin, E. Lamboray, and M. Gross, "3D video fragments: dynamic point samples for real-time free-viewpoint video", Computers and Graphics 28 (1), Special Issue on Coding, Compression and Streaming Techniques for 3D and Multimedia Data, pp. 3-14, Elsevier Ltd, 2004.

[44] E. Lamboray, S. Wümlin, M. Waschbüch, M. Gross, and H. Pfister, "Unconstrained Free-Viewpoint Video Coding", Proceedings of the IEEE International Conference on Image Processing (ICIP) 2004, Singapore, October 24-27, 2004.

[45] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on Multi-view Video Coding", Doc. N7327, Poznan, Poland, July 2005.

[46] ISO/IEC JTC1/SC29/WG11, "Requirements on Multi-view Video Coding v.2", Doc. N7282, Poznan, Poland, July 2005.

[47] S. B. Kang, R. Szeliski and P. Anadan, "The Geometry-Image Representation Tradeoff for Rendering", Proceedings of International Conference on Image Processing, Vol. 2, pp. 13-16, 2000.

[48] C. Zhang, T. Chen, "A survey on image based rendering – representation, sampling, and compression", Signal Processing: Image Communication 19, pp. 1-28, 2004.

[49] M. Magnor, B. Girod, "Data Compression for Light Field Rendering", IEEE Trans. On Circuits and Systems for Video Technology, Vol. 10, No. 3, pp. 338.343, April 2000.

[50] G. Alregib, Y. Altunbasak, J. Rossignac, "Error-resilient transmission of 3D models," ACM Transactions on Graphics (TOG), Volume 24 , Issue 2, Pages: 182 - 208, April 2005.

[51] B. Tari, Y. Yemez, Ö. Özkasap, R. Civanlar, "Progressive View-Dependent Transmission of 3D Models over Lossy Networks", Proceedings, EUSIPCO-2005, Antalya, Sept. 2005.

[52] M. Kunter , S. Knorr, C. Clemens, and T. Sikora, "A gradient based approach for stereoscopic error concealment", Proc. ICIP 2004, IEEE International Conference on Image Processing, Oct. 2004.

[53] E. Kurutepe, M. R. Civanlar, A. Murat Tekalp, "A Receiver-Driven Multicasting Framework For 3DTV Transmission", Proceedings, EUSIPCO-2005, Antalya, Sept. 2005.

[54] TC Poon, Recent Advances in Optical Scanning Holography, Journal of Holography and Speckle, 1, 6-25 (2004)

[55] T Ito, Color Electro-holography by Three Colored Reference Lights Simultaneously Incident on One Hologram Panel, Optics Express, 12, 18 (2004)

[56] T Kreiss, P Aswendt and R Hofling, Hologram Reconstruction using a Micro-mirror Device, Opt Eng 40, 926-933 (2001)

[57] GE Favalora *et al*, Volumatric Three-dimensional Display System with Rasterization Hardware, Proc SPIE 4297, 227-235 (2001)

[58] K Langhans *et al*, FELIX 3D Display: an Interactive Tool for Volumetric Imaging, Proc SPIE 4660, 176-190 (2002)

[59] I Sexton, Parallax Barrier Display Systems, IEE Colloquium: Stereoscopic Television, Digest No. 1992/173, 5/1-5/5 (1992)

[60] C Moller and ARL Travis, Flat Panel Time Multiplexed Autostereoscopic Display Using an Optical Wedge Waveguide, Proc 11th International Display Workshop, Niigata, Japan, 1443-1446 (2004)

[61] ST Zwart *et al,* A 20-in Switchable Autostereocopic 2D/3D Display, Proc 11th International Display Workshop, Niigata, Japan, 1459-1460 (2004)

[62] S-S Kim *et al*, The Optical Design and Analysis for Super Multiview Three-dimensional Imaging System, Proc SPIE 4297, 222-226 (2001)

[63] H Urey, Retinal Scanning Displays, Encyclopedia of Optical Engineering, R Driggers (Ed), 3, 2445-2457 (Dekker, 2003)

[64] L. Onural and H. M. Ozaktas, Signal Processing Issues in Diffraction and Holographic 3DTV, Proceedings, EUSIPCO-2005, Antalya, Sept. 2005.

[65] L. Onural, Sampling of the Diffraction Field, Applied Optics, vol 31, no 32, pp 5929-5935, Nov. 2000.

[66] H. M. Ozaktas, Z. Zalevsky and M. A. Kutay, The Fractional Fourier Transform with Applications in Optics and Signal Processing, Wiley, New York, 2001.

[67] L. Onural, T. Sikora, and A. Smolic: "An Overview of a New European Consortium: Integrated Three-Dimensional Television - Capture, Transmission and Display (3DTV)", Proceedings of European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT), 2004.

[68] http://www.3dtv-research.org