Levent Onural

Signal Processing and 3DTV

Frequently, the most important sporting events are used as platforms to showcase and sometimes launch the latest communication technologies. Over the years, we have enjoyed watching the Olympic Games or the World Cup soccer games with new technologies such as satellite broadcasting, color TV, and high-definition TV. The 2010 World Cup continues this tradition by broadcasting the games in three-dimensional (3D) TV (3DTV) [1]. This article presents an introduction to the technological issues facing a broad deployment of 3DTV systems and discusses some of the signal processing techniques that are used or need to be developed in this area.

WHICH 3D?

One of the issues in 3DTV is the definition of what 3D is, or what really 3D means. Usually this question is asked to draw attention to the fact that recognized standardization bodies have not yet defined a common operating format that can be referred as the 3DTV; maybe the question itself is an expression of dissatisfaction with this situation.

However, it is quite straightforward to define the ultimate 3DTV, at least in the physical sense: "Seeing" is a consequence of optical interaction of the observer's light sensors with the light that reaches to them. If the same volume-filling light distribution with all relevant physical properties can be generated, at a remote location and maybe at a different time, then the observer that interacts with the same light will see the same image with all its 3D features as if looking at the original. Therefore, a ghost-like optical duplicate will be observed [2]. Consequently, for

an ideal 3DTV system design, one should seek i) to understand the light with all its physical properties and ii) should develop techniques to capture, process, and recreate the same light in a volume of observation. Such an ideal imaging mode as described above is called "true 3D." Holography and integral imaging are two well-known techniques that target capture and replay of physical light fields. Such a physical duplication of light will create 3D not only for humans but for all sorts of other living or camera-type observers.

Yet what is known today commonly as 3DTV is not even close to the ideal case described above. Instead, due to technological limitations, what is launched many times with different degrees of quality throughout the history as 3DTV is based primarily on a different approach: it relies on human perception and how the human visual system processes incoming optical stimulus. For example, stereoscopy has been known for more than 170 years and is based on delivering two appropriate two-dimensional (2D) images (video for the TV case) to the two eyes of the observer. By the way, stereoscopic TV was invented, only a year after its 2D counterpart, in 1929 [6]. So for many, 3DTV means simply two 2D views of a scene captured from slightly two different angles that match the locations of the two eyes of a human observer. This kind of low-end 3DTV cannot deliver true 3D experiences.

And then there are many different 3DTV variants between the most sophisticated ultrarealistic ideal case outlined above, and the simplest stereoscopic TV mentioned afterwards: for example, in multiview video, many 2D videos of the same scene are shot, each from a different angle, and thus a different video pair can be delivered to the
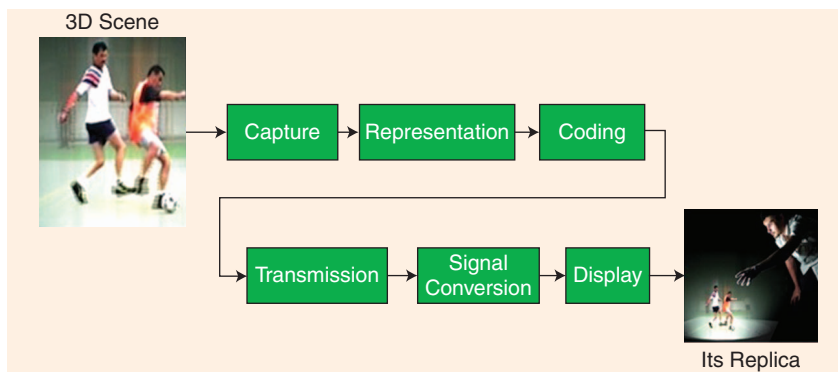
human observer as he changes his observation location, or to support multiple viewers at different locations.

Therefore, as in many other commercialization stories of novel technologies, attempts to deliver consumer goods are made starting from those variants that can be handled by the available underlying technologies; these technologies are primarily optics and electronics, together with telecommunications, signal processing, and others, for TV.

With these observations, it is quite easy to see that short-term efforts in 3DTV will be based essentially on simple stereoscopy, followed by medium-term activities that focus on multiview video. Only long-term research will drive the ultrarealistic approaches like holography and integral imaging. Being among the later stage activities that follow essential research and development phases, standardization, and commercialization efforts will also follow this time line. Currently, emerging standards cover only simple stereoscopy and some low-end multiview video. Therefore, in terms of research, we see concurrent activities that encompass early phase high-end true 3D targeted research, together with currently main-line multiview video research and development and final-stage development, which is targeted primarily to improve stereoscopic techniques for consumer products. Therefore, research in 3DTV will continue for many more years, shifting the research momentum to newer true 3D technologies as the stereoscopic and multiview video techniques mature [5].

3DTV CHAIN

Even though what the consumer interacts with is almost always the display

3D Scene

Capture → Representation → Coding

Transmission → Signal Conversion → Display

Its Replica

[FIG1] Functional units of a possible end-to-end 3DTV chain (reused from [7]).

device, the TV involves a large chain of processes that starts with video capture and continues with an intermediate representation, coding and compression, transmission, display side processing, and, eventually, the display stages. Each such stage, which collectively form the 3DTV, may be implemented based on a number of quite different techniques [3].

For example, it is possible to capture 3D content by a single 2D camera, but it is quite likely that two or more 2D cameras will be used in most 3DTV systems. However, it is possible to get help from projected structured light, for example, projecting periodic stripes onto a 3D object will create pictures that carry the curvature information in the form of distorted stripes. Alternate techniques include time-of-flight cameras for depth data recording. Such cameras generate short duration light pulses with a planar shape, like a curtain made up of light; upon hitting the 3D object, reflections from different depths send back a light pulse that is no longer planar but rather looks like the mask of the object. This information is captured by the camera that processes the reflected pulse. Holographic cameras have also been demonstrated. The principle of a holographic camera is to illuminate the object with a coherent light and then record the interference of the reflected light by a reference beam; coherent light sources are the lasers. The interference pattern carries the complete 3D information.

There is no doubt that what is actually captured is determined by considering what is needed by the display, and vice versa. However, the level of such coupling differs with each design. Simpler designs will end up having a very tight coupling and therefore only a specific form of a display can be driven by a specific form of a camera setup. At the other extreme, there will be a complete decoupling of the input and display via an intermediate abstract representation that is constructed by captured data and then utilized by different types of displays. Therefore, the representation stage right after the input capture may be simply "do nothing" operation in low-end approaches, or a quite complicated step to generate, for example, complete time-varying 3D scene models in sophisticated 3DTV systems. Such high-end systems may use intermediate point-cloud representations, or even 3D time-varying meshes with texture. In point-cloud representations, a 3D object is represented by a dense set of point samples taken over the object; each point has a 3D coordinate and a color data attached to it. Wire-mesh models are commonly used in computer graphics: simply a curved surface is approximated by planar patches that are usually triangular; a wire-mesh object is then converted to a photorealistic object by sticking the color variation (i.e., the "texture") over the surface.

Since 3D video data is highly redundant, the captured data is highly compressible, no matter which representation form is employed. In addition to adapting common lossy and lossless compression techniques to 3D video, one can exploit the statistical nature of 3D video to come up with novel more specific compression techniques. Such work has been reported [3].

The transport of 3D video content has its own specific problems and nature, but well-known 2D video transport techniques can be modified to be used for 3D as well. In particular, error concealment techniques to compensate for lost data is quite different in different forms of 3D video delivery [3].

The display of 3D content can take many quite different forms: while displays whose bulk appearance is similar to 2DTV sets are getting quite popular for consumer applications, volumetric displays have also been used for many specific purposes. More advanced displays based on some limited forms of light field rendering techniques are commercially available. Prototypes of some limited forms of holographic displays have been demonstrated [3].

SIGNAL PROCESSING ISSUES IN 3DTV

The basic features of the 3DTV chain as presented above imply that signal processing is needed at all stages and that includes signal processing right in the cameras and in the display devices. However, the two stages where signal processing issues dominate are right after the capture at the transmitting side and right before the display at the receiving side, in the form of interfaces. Furthermore, the compression of captured data is another intrinsically signal processing intensive stage.

Capture-side signal processing issues include primarily different forms of data fusion operations. In the simplest case, the data captured by two cameras in stereoscopic TV must be processed to correct alignment problems and color mismatches. More complicated stereoscopic video processing involves disparity modifications for proper parallax and associated perceived depth alterations. Resolution modifications are also common. More sophisticated capture units that provide multiview video or video-plus-depth data naturally bring more sophisticated subsequent signal processing needs. A few typical examples are color and geometry corrections among different 2D video sources, virtual camera techniques to interpolate and generate 2D views from intermediate

angles between physical camera locations, and filling missing data as occlusions change due to changing viewing angles. But the ultimate signal processing that caters for sophisticated future 3DTV operations is the construction of a complete time-varying 3D model, in the form of point clouds, meshes, or their variants, by fusing data from a multitude of input devices including cameras and sensors. If the captured data is in holographic form, the subsequent signal processing is quite complicated and specific to associated optical wave propagation, diffraction, and interference phenomena.

Different compression techniques are used for different forms of data that emerge after processing as described above. Simpler ones employ 2D video compression techniques by shuffling frames from multiple 2D video cameras to obtain a single sequence. More sophisticated techniques utilize the redundancy in frames from different 2D cameras, together with temporal redundancies, more directly; quite complicated referencing techniques among multivideo frames are proposed and utilized. Furthermore, compression techniques for complicated time-varying 3D scene models are also investigated. As expected, the nature of holographic data is totally different than classical video frames and, therefore, justifies specific intraframe coding techniques of its own. In any case, it should not be forgotten that the original 3D scene is naturally highly redundant, and therefore any successful 3D video compression procedure should generate a reasonable bandwidth which is somewhat more than a single 2D video but not prohibitively large.

Display-side signal processing is much more demanding compared to classical 2DTV. The main reason is the fact that 3DTV displays may have totally different physical structures. Some are pixelated devices tightly coupled with optics such as lenticular screens for autostereoscopic viewing. Some sophisticated ones target to generate light fields that then give the 3D output that can also be observed without any special glasses. A classical 2DTV display writes the frames on its screen in a raster-scan fashion. However, what is written on a 3D display could be totally different. For example, in lenticular autostereoscopic displays, a compound single frame is generated by mingling many video frames from different cameras by using a technique called "interzigging." Even in the simplest, time-multiplexed stereoscopic displays with goggles, the refresh rates and raster structures are quite different and more demanding than the 2D case. Therefore, the interface to convert received data to the specific display type is not a trivial task; successful signal processing applications will definitely make a lot of difference in terms of speed, efficiency, and perceived 3D quality. The most demanding signal processing needs arise in the case of holographic displays: for example, conversion of a given 3D model into holographic fringe patterns needed by a holographic display has the potential to push the limits of existing signal processing techniques, and thus can start a completely new line of fundamental processing modes and underlying mathematics [4]. Signal processing techniques for holographic 3DTV are based on optical wave propagation fundamentals and closely linked to signal processing concepts like Fourier decompositions (plane waves are 3D Fourier basis functions), sophisticated sampling and recovery techniques other than commonly used Shannon case, convolutions with strange kernels, difficult inverse problems, decompositions using unusual basis functions, etc. For example, the relationship between the field patterns over two parallel planes in space due to coherent monochromatic optical wave propagation is the starting point in holography, and such a relationship is correctly modeled as a linear shift invariant system [4]. When the commonly used Fresnel approximation is employed, the convolution kernel of the linear shift invariant system that represents the field relation between the parallel planes is quite interesting: the kernel is neither time limited nor band limited. The problem becomes difficult when one of the planes become an arbitrary surface to represent a 3D object, because, although the relationship is still linear, it is not shift invariant.

A challenging signal problem is the automated conversion of existing 2D content to 3D. It is quite unlikely to fully automate such a process; however, quite successful semiautomatic, somewhat supervised techniques are already demonstrated.

A major problem in low-end 3DTV (and in 3D cinema) is the potential viewer discomfort, which is a motion sickness type of a feeling; it is usually called eye fatigue. This is a result of conflicting perception cues to the brain, and the conflicting perception cues are intrinsic in systems based on stereoscopy; they get worse if the stereoscopic pair is not presented right. Even though such problems cannot be completely eliminated in such systems, signal processing can remove most fundamental sources like alignment mismatches, and therefore are key to commercially successful end results. Squeezing depth variations is another technique which reduces eye fatigue at the expense of reduced 3D experience. Complete removal of this severe problem is possible only by providing true 3D displays as outlined above.

There is no doubt that the signal processing community will enjoy 3DTV-related topics, which are quickly gaining popularity.

AUTHOR

Levent Onural (onural@ee.bilkent.edu.tr) is a professor at Bilkent University, Ankara, Turkey.

REFERENCES
[1] ipTV News. (2010, May 6). World Cup to kick start uptake of 3DTV [Online]. Available: http://www.iptv-news.com/iptv_news/may_2010/world_cup_to_kick_start_uptake_of_3dtv

[2] L. Onural, "Television in 3D: What are the prospects," *Proc. IEEE,* vol. 95, no. 6, pp. 1143-1145, June 2007.

[3] Special Section on 3DTV (A Collection of Six Papers and an Introduction), *IEEE Trans. Circuits Syst. Video Technol.,* vol. 17, no. 11, pp. 1566–1658, Nov. 2007.

[4] L. Onural and H. Ozaktas, "Signal processing issues in diffraction and holographic 3DTV," *Signal Processing: Image Commun.,* vol. 22, no. 2, pp. 169–177, Feb. 2007.

[5] H. M. Ozaktas and L. Onural, Eds. *Three-Dimensional Television—Capture, Transmission, Display.* New York: Springer-Verlag, 2008.

[6] O. Schreer, P. Kauff, and T. Sikora, Eds., *3D Videocommunication: Algorithms, Concepts, and Real-Time Systems in Human Centric Communication.* Hoboken, NJ: Wiley, 2005.

[7] L. Onural, H. M. Ozaktas, E. Stoykova, A. Gotchev, and J. Watson, "An overview of the holography related tasks within the European 3DTV project," in *Proc. SPIE,* vol. 6187, 2006, pp. 61870T–1–61870T–10. **[SP]**