Digital 3DTV

Alexey Polyakov

The advent of the digital 3DTV era is a *fait accompli*. The question is: how is it going to develop in the future? Currently the Standard definition(SD) TV is being changed into High definition(HD) TV. As you know, quantitative changes tend to transform into qualitative ones. Many observers believe that the next quantum leap will be the emergence of 3D TV. They predict that such TV will appear within a decade. In this article I will explain how it is possible to create stereoscopic video systems using commercial devices.

Brief historical retrospective

The following important stages can be singled out in the history of TV and digital video development:

1—**black and white TV** – image brightness is transmitted.

2 – **colored TV** – image brightness and color components are transmitted. From the data volume perspective, adding color is a quantitative change. From the viewer perspective, it is a qualitative change.

3 – **digital video emergence** (Video CD, DVD) – a qualitative change from the data format perspective.

4 – **HD digital video and TV** (Blu-Ray, HDTV) – from the data volume perspectiveit is a quantitative change. However, exactly the same components are transmitted: brightness and color.

Specialists and viewers have been anticipating for the change that had been predicted by sci-fi writers long ago, - **3D TV emergency**. For a long time data volume was a bottleneck preventing stereoscopic video demonstration as the existing media couldn't transmit it. Digital TV enabled to transmit enough data and became the basis for a number of devices that helped to realize 3D visualization.

How should stereoscopic TV look like?

Let's state the main requirements for 3D TV being used widely in the domestic conditions.

From the user (viewer) perspective:

1. The reproducing device should create a realistic effect of a 3D image.
2. The picture should be watched naturally without any tension. No additional devices should be required (i.e. helmet or special glasses). It should be possible to watch either by 1 or several viewers at the same time.
3. The display device should be able to show both stereoscopic and conventional 2D images.
4. The display device should be small enough and convenient for home usage.

From the designer standpoint, general requirements for 3D TV media and technical devices are the following:

1. Data volume that is required to display a stereoscopic image can't considerably exceed data volume transmitted for displaying a conventional image.
2. The transmission method should be compatible with the existing standards and technologies.

It seems very difficult to satisfy all these demands at the same time. However, the complex of modern technical solutions makes it possible.

Physical principles of the stereoscopic perception

Human beings perceive the world in 3D due to a number of phenomena: geometrical and aerial perspective, shadows and flares on object surfaces, relative sizes of objects. Graphic methods modeling such phenomena are used by painters from long ago to depict the 3D nature of objects drawn on the flat.

Nature gave human beings binocular vision – a pair of eyes placed at the distance of 60-70mm from each other, which enables to see the world from 2 observing points simultaneously. As a result, the images seen by the right and the left eye are a little bit different. These two images are called "a stereo pair". Analyzing the differences between the images in a stereo pair the human brain gets the information about dimensions and distances to the objects being viewed (Fig. 1).

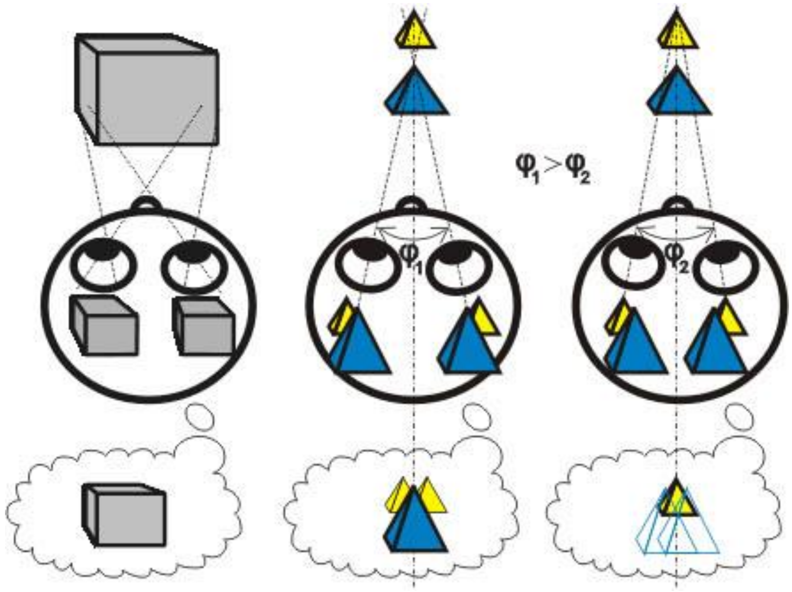

Fig. 1. Parallax: each eye sees an object in its own way; the brain estimates the difference and forms a 3D image (source - www.triaxes.ru).

The apparent shift of an object as a result of a changed view point is called "parallax". It is the main factor in the process of 3D world perception.

All methods that are widely used to create a stereo video effect use the separate view principle – the left eye sees the left image of a stereo pair, the right eye – the right one. The difference lies in the way the separation of a stereo pair is carried out. Most of the modern stereo display devices, either in the cinema or TV, are based on methods known for more than a hundred years.

Anaglyph display method is 150 years old. The method was proposed by D'Almeida and Ducos du Hauron in 1858. It was realized in the cinema by Louis Lumière in 1935. The anaglyph method (Greek "anagliphos" – texturized) consists in coloring the images of a stereo pair into additional colors. Both frames form 1 image. Special glasses colored in corresponding colors help to separate the left and the right frame. The anaglyph method is used in the movie screening and TV broadcasting. The advantage of this method lies in its simplicity and the low price. Among the disadvantages – partial color loss and the necessity to use glasses.

The method of polarization is 120 years old. It was proposed by J. Anderton in 1891. It became widely spread after the invention of polarization foil by E. Land in 1935. The left and the right frame are projected simultaneously but the light is polarized (linearly or circularly) in different directions. The image is viewed with glasses that have corresponding light filters. The method of polarization became widely spread in the cinema screening due to the accurate separation of a stereo pair and color retention; the disadvantages of this method are the necessity to use expensive equipment, special display devises and glasses that a viewer must wear. The method is used in IMAX cinemas, etc.

Lenticular stereo is 110 years old. The glasses-free stereo method using parallel light absorbing lenticular material was introduced in 1896 by Berthier and Liesegang simultaneously. For the first time in the world this method was introduced by S. Ivanov and A. Andrievskiy in the USSR in 1942 to demonstrate stereo movies and was carried out under the supervision of B. Ivanov. The first cinema with the lenticular screen in the world calles "Stereokino" appeared in Moscow in 1947. The lenticular screen consisted of a number of opaque vertical strips. The light went through transparent areas between strips and each eye of a spectator saw the necessary fragment of the image. The size of the screen was 3x3 m.

Such devices of glasses-free 3D visualization are called "**stereoscopic"**. This method has various functional realizations: lenticular and barrier. Now the lenticular method of the display design is chiefly used. To show the content using the lenticular method, a source stereopair is "cut" into vertical strips that interchange so that under each lens there's a pair of strips: one - from the left frame, the other – from the right one. Such "striped" image is called "encoded". The working principle of the lenticular screen is shown in Fig. 2. The stream of light emanating from an encoded image goes through lenses and splits up so that the left eye of a viewer sees the left image, and the right eye sees the right one.

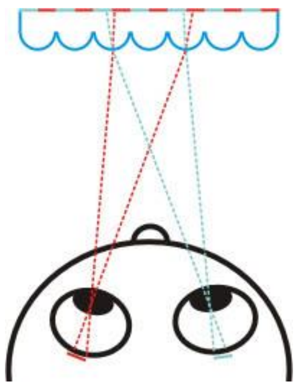

Fig. 2. The scheme shows the way the encoded image of a stereopair is splitted with the help of the lenticular material. The left image is marked by the red color, the right one – by the blue color.

The effect of the lenticular method is greater when, instead of 2 frames of a stereo pair, a series of frames with a slight horizontal shift (multiview shooting) is shown. In this case there occurs a wide stereovision area. As a result a viewer can move watching the scene from different angles and it seems possible to look behind the foreground objects. It makes the stereo image look natural. When shooting series of frames photographers use special stereo cameras with a number of lenses (Fig.3), or special tripods allowing to move the camera horizontally while shooting.

Fig.3 Multi-lens stereo camera (source - www.3-dimages.com)

Fig. 4 Triaxes StereoRail tripod for stereo multiview photo and video shooting (source - www.triaxes.ru).

The advantage of the lenticular method is that the separation device is combined with the image therefore the viewer doesn’t have to wear any glasses. Moreover, a 3D image, formed out of a series of frames shot from different angles, makes the scene look more real.

The disadvantage is that it requires much more data to display a 3D image of high quality. For the anaglyph and polarization methods 2 frames of a stereo pair are enough. For the lenticular method it is preferable to display 9-12 frames simultaneously. I am going to tell how to solve this problem hereafter.

Stereoscopic displays (TV-sets)

There are a lot of types of stereoscopic monitors. Almost all famous brands (LG, Philips, Sharp, etc.) have display models based on the lenticular principle. It can be explained by the fact that such devices meet the 5 above mentioned requirements. The development of the lenticular stereo monitors began in the previous century (Philips). But a really good effect and commercial success have been achieved relatively recently when the HD standard became wide-spread. It is connected with the encoded multiview image that requires higher resolution than each of the source frames. Only with the ability to transmit and decode HD video the quality of the 3D effect increased as the amount of pixels was sufficient now.

2D + Z format

Obviously, multiview shooting is not an easy task. You need either a special camera with a big number of lenses or a series of cameras and a device enabling synchronous shooting. You should also keep in mind that you will have to store a lot of data – video streams from each camera. Even though modern approaches of the digital video compression enable to take into account time and spatial redundancy, the data volume increases many times when you perform multiview shooting.

One of the effective methods to solve the data problem is using the so-called 2D+Z format. Any conventional (2D) image can be supplied with the information about the distance from the viewer to each pixel (Z-coordinate). Such image representation is called “the 2D+Z format” while the Z coordinate plane is called “the depth map”. It can be presented in the form of a monochrome image. In the depth map the distances to the dots from a viewer are represented by the grey gradations. In Fig. 5 there is an example of a source mage and its depth map.

Fig 5. Source mage and its depth map (source - www.stamptex.pl)

The 2D+Z format continues the conception of displaying information by components. As it is known, both in the analogue and digital television the image is formed from the brightness of 2 color elements. Adding one more element, which characterizes dimensions of the image, seems quite a logical development and conforms with the principles of compatibility. In fact nowadays black and white TV-sets work successfully using only the brightness component of the TV image, while colored TV-sets use all the data. In case the video containing the information on the image dimension is broadcasted, additional data can be used by the display devices that can interpret them correctly and ignored by the others. Using 2D+Z enables to broadcast stereoscopic video with the data flow increased by 25-30% only. Thus, the requirement to keep the acceptable data volume is satisfied.

Re-creating a multiview image

2D+Z is not, however, a multiview series or a stereo pair. To show a 3D image it is necessary to calculate a frame series. Stereoscopic image creation is achieved through source image interpolation with the depth map taken into account. Then the final frame series is demonstrated with the help of a lenticular display (Fig. 6).

Fig. 6. 2D+Z image transformation into a frame series and its demonstration on a 3D display (source - www.philips.com, www.isu3d.org).

Standardization

The MPEG-2 format has an additional multiview profile accepted in 1996 that enables to encode and transmit an image from 2 and more cameras (see JENS-RAINER OHM. STEREO/MULTIVIEW VIDEO ENCODING USING THE MPEG FAMILY OF STANDARDS). MPEG-4 format (part 2) specification determining the encoding method enables to transmit a conventional 2D image as well as its depth map (Z) using the standard procedure. MPEG-4 (part 10/AVC) provides analogous capability. Moreover, the standards enable to encode the Z plane as additional data that can be ignored by the devices that are not supposed to deal with it. Thus, Z presence will not influence the decoding of the main image. The MPEG-2 and MPEG-4 (AVC) formats are the main formats in digital TV that's why there exists a sufficient basis for standard 3D video content broadcasting. It should be noted that this standardization process is going on and we expect the new additions in 2011 (see ATANAS GOTCHEV. COMPUTER TECHNOLOGIES FOR 3D VIDEO DELIVERY FOR HOME ENTERTAINMENT.].

Technical realization

Current technologies enable to begin immediate realization of the stereoscopic broadcasting systems on the internet (IPTV). In computer networks it is possible to provide 3D services individually instead of broadcasting the video stream for thousands of viewers while only several percent of them have 3D monitors. A number of world display manufacturers produce models enabling to display 3D images. In particular, Philips manufactures 3D monitors based on the lenticular principle. The input data for such monitors are a conventional image plus a depth map (2D+Z format) [3D INTERFACE SPECIFICATIONS. WHITE PAPER. PHILIPS 3D SOLUTIONS]. Such monitors became wide spread and are successfully used to demonstrate adds in huge trade show centers. It is informed that they are used in test 3D IPTV systems [SEE PHILIPS AND EVENTIS DEMONSTRATE 3D VIDEO-ON-DEMAND AT IBC2007]. The disadvantage is that it is necessary to process the image before displaying it on a monitor. Nowadays the processing is done with the help of a computer and special software. Using a computer is not always easy and limits the application field of such monitors.

As it was noted above, to transmit 2D+Z video data we can compress it in accordance with the common MPEG-2 or AVC standards. Standard streams can be decoded with small digital video decoding devices -set top boxes (STB). However, STBs produced by many world manufacturers do not generate the necessary control information for the correct interpretation of the depth data. The Russian company Elecard produces a series of STBs based on Philips Nexperia and TI DaVinci DSP-processors [ELECARD ITELEC STB 61x]. The peculiarity of these devices is that they combine the features of a hardware decoder (small size, low power drain (about 6-8 W), reliability) and software flexibility (the functionality can be easily enhanced). The architecture of these devices enables to change the embedded software. It allows to enhance the functionality easily without changing the hardware component. For example, it makes possible to predict the ability to work with 3D images for a batch of devices.

There's a special software version developed for STBs that forms special information that is required to display video in the 2D+Z format on a 3D monitor.

The working prototype of the device is shown in Fig. 7. The picture shows the real-time decoding of the MPEG-2 1920x1080 (HD) stream in the 2D+Z format. The whole picture is 1920 wide but on the monitor only the left part is shown. The right part is depth information.

Fig. 7 Elecard 3DSTB prototype. MPEG-2 1080 HD video stream decoding in the 2D+Z format.

3D video content preparation

3D TV system creation is impossible without the sufficient amount of 3D video materials, means of conversion into various formats and the appropriate codecs. Many world film industries have already begun to shoot new movies in the stereo format, i.e. using 2 cameras at once (Fig. 8).

Fig. 8. Camera for 3D video shooting (source - www.inition.co.uk).

Such shooting allows to demonstrate stereoscopic films in the cinemas and on monitors based on the polarization technology. Although 2 frames are not enough to demonstrate 3D video on stereoscopic devices. As it was noted above, the 2D+Z format is the most appropriate format due to the data volume being transmitted and the ability to recreate a stereo image. That's why the task solution to transform a stereo pair into the 2D+Z format seems up-to-date. Mathematically this task means finding relative distance to the objects using 2 images. There have already appeared software products enabling to calculate Z (depth map). For example, TRIAXES STEREOTRACER allows to calculate a depth map for a stereo pair of a photograph automatically, BLUEBOX FROM PHILIPS 3D SOLUTIONS does the same with the video content.

Despite the fact that there already exist means to calculate 2D+Z, this field brings vast opportunities for research and development of new products.

One more difficult task is to transform a great amount of 2D materials into the 3D format. A lot of world companies are involved in such development as well as our company - Triaxes Vision, which is based on 2 companies Triaxes and Elecard and aimed to develop 3DTV systems.

A project introduced by Triaxes Vision for the Enterprise Development Support Foundation in the technical-scientific sphere won the first place in the Siberian Federal District according to the experts' research.

The project includes the following task solving:

- Mathematical algorithm and 2D→3D(2D+Z) video conversion software development
- 3D data encoding into the required format
- Assuring compatibility of the 3D video conversion formats with the digital TV broadcasting equipment being used nowadays and the one being planned for the 2015.
- Designing the transmitting and decoding scheme at the receiving end;
- Displaying 3D video on the monitor (TV-set);

Now in the framework of the project we have already developed the modifications of the standard MPEG-2 and MPEG-4 (AVC) video codecs used to compress 3D video data. We have also developed modules transforming video into 2D+Z format as well as a program unit for 3DSTB creation.

3D IPTV system prototype

Elecard and Triaxes Vision companies created a prototype of a system delivering 3D video on demand. The system includes a server with a 3D movie library, which provides the VOD (video on demand) service, a distributed network of viewers and stereo monitors equipped with 3D STBs (Fig.9).

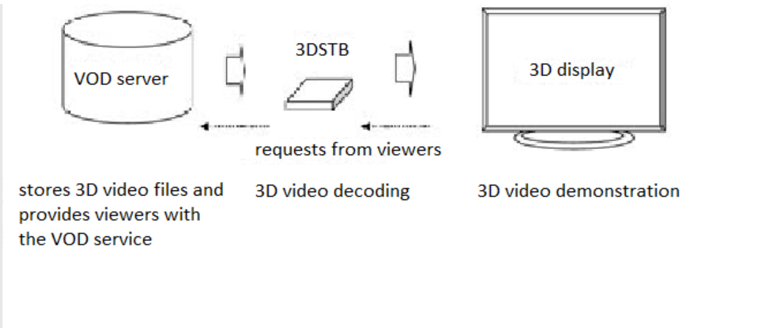

Fig. 9. Scheme of a 3D IPTV system prototype

The application fields of such systems can be 3D video broadcasting in the IPTV networks for individual viewers as well as digital video delivery on big screens to demonstrate advertising information in the stereoscopic format.