Vol 2 | ISSN 2041-5923
2010

IET Journals

The Best of IET and IBC

INSIDE Papers and articles on electronic media technology from IBC 2010 presented with selected papers from the IET's flagship publication *Electronics Letters*.

Published by the Institution of **Engineering and Technology**

Contents

Selected content from IBC 2010

Selected content from the IET

ISSN 2041-5923

Introduction

Welcome to The Best of IET and IBC 2010, a joint publication of the International Broadcasting Convention and the Institution of Engineering and Technology.

As a member of the IBC partnership board, the IET maintains a close working relationship with IBC and, following the success of our first joint publication at IBC 2009, we are pleased to bring you the 2010 volume of The Best of IET and IBC. This publication will appeal to everyone with a technical interest in the field of broadcast media, containing the very best technical content from IBC 2010 combined with relevant selected content from the IET's flagship peer-reviewed journal *Electronics Letters*.

The IBC 2010 content includes a selection of papers from across the sessions as well as the selected Best Paper: Dr. K. Murray's 'Does size matter? The impact of screen size on stereoscopic 3DTV'. Inline with the IBC's focus on young professionals, this volume also includes the two best poster papers contributed by young professionals, alongside an interview with the selected best young professional poster author, Gustavo Marra. All this is complemented by specially selected recent papers from *Electronics Letters*, the IET's unique broad-spectrum, rapid publication journal, as well as an *Electronics Letters* interview with David Wood of the European Broadcasting Union, giving some background to his invited Insight Letter for *Electronics Letters* also published in this volume.

As always IBC is committed to staging the world's best event for professionals involved in content creation, management and delivery for multimedia and entertainment services. IBC's key values are quality, efficiency, innovation, and respect for the industry it serves. IBC brings the industry together in a professional and supportive environment to learn, discuss and promote current and future developments that are shaping the media world through a highly respected peer-reviewed conference and a comprehensive exhibition, plus demonstrations of cutting edge and disruptive technologies. In particular, the IBC conference offers delegates an exciting range of events and networking opportunities, to stimulate new business and momentum in our industry. The spotlight this year is on the creativity that we see being brought back into the marketplace and how innovation and enterprise are changing the landscape in the communications and media industries. The IBC 2010 conference committee continues to craft an engaging programme in response to a strong message from the industry that this is an exciting period for revolutionary technologies and evolving business models.»

Left: The IBC 2009 Innovation Awards

Right: The launch event for The Best of IET and IBC 2009

The IET is one of the world's leading professional societies for the engineering and technology community, with more than 150,000 members in 127 countries and offices in Europe, North America and Asia-Pacific. It is also a publisher whose portfolio includes a suite of 22 internationally renowned peer-reviewed journals covering the entire spectrum of electronic and electrical engineering and technology. Many of the innovative products that find their way into the exhibition halls of IBC will have originated from research published in IET titles with more than a third of the IET journals covering topics relevant to the IBC community (e.g. *IET: Image Processing*; *Computer Vision*; *Communications*; *Information Security*; *Microwave Antennas & Propagation*; *Optoelectronics*, *Circuits & Systems* and *Signal Processing*). The IET papers contained in this publication come from the IET's flagship journal, *Electronics Letters*, which embraces all aspects of electronic engineering and technology. *Electronics Letters* has a unique nature, combining a wide interdisciplinary readership with a short paper format and very rapid publication; produced in print and online fortnightly. Many authors choose to publish their preliminary results in *Electronics Letters* even before presenting their results at conference, because of the journal's reputation for quality and speed. This year *Electronics Letters* has a fresh new look, bringing its readers even more information about the research through a contemporary colour section that includes author interviews and feature articles expanding on selected work from each issue.

Working closely with the IET journals team are the IET Technical and Professional Networks. The TPNs exist to act as a natural home for people who share a common interest in a topic area (regardless of geography); foster a community feeling of belonging and support dialogue between network registrants, the IET and each other. Assisting each TPN is an executive team, made up of willing volunteers from that network who bring together their unique experience and expertise for the benefit of network registrants. Members of the Multimedia Communications Network[1] executive team have played an essential role in the creation of this publication in reviewing, suggesting and helping to select content. They have contributed their industry perspectives and understanding to ensure a relevant and insightful publication for the broad community represented at this year's conference, showing the key part volunteers have to play in developing the reach and influence of the IET in its aim to share and advance knowledge throughout the global science, engineering and technology community.

Finally, we would like to extend our thanks to everyone involved in creating the 2010 edition of The Best of IET and IBC. We are sure you will find it indicative of this year's exciting conference programme as well as the high quality peer-reviewed research published by the IET. We hope you will enjoy it, and wish all of you attending this year an enjoyable IBC 2010.

Michael Lumley
Chairman of the IBC Conference
&
The IET Multimedia Communications Network executive team

[1] www.theiet.org/multimedia

ISSN 2041-5923

Editorial

IBC's crème de la crème of conference contributions

In early February of each year IBC closes its online call for conference contributions and with great anticipation we download between 250 and 300 synopses from eager potential authors. The synopses originate from every corner of the globe and not only span the vast range of modern media technologies, but occasionally provide exciting glimpses of entirely new engineering concepts or systems which are being implemented on a breathtakingly ambitious scale.

The busy conference schedule means, however, that there will only be opportunities for perhaps 42 authors to present their work to session audiences, so IBC's Technical Papers Committee must decide which synopsis authors will be invited to produce full conference papers. Our 16 members from across the media industry first grade the quality of all the submissions and then categorise them into session areas, which reflect the year's most topical industry interests. Not surprisingly, some areas are more heavily represented than others; this year 3D technologies dominated and competition to author papers in this session was especially keen.

In choosing papers, our primary concern is with novelty; we look for contributions which provide new and relevant results, solutions or ideas, which will be of strategic interest to the media industry. Of course, we also expect papers to state their arguments clearly and provide a balance of background, technology, results and conclusions. We have an eye too, for papers which are likely to provide entertaining presentations and stimulating discussion – an important element of design always goes into a successful conference session.

At IBC we also value our poster sessions, regarding them with the same status as our presentational sessions. All poster authors may also submit a full conference paper for inclusion in the published proceedings and these are peer-reviewed with the same rigour as all the others. We regard the poster much like a small (but non-commercial) exhibition stand, providing an opportunity for its authors to interact with interested delegates in a far closer manner than would be possible in a formal presentation. Small demonstrations can also be given. Many large companies now choose the poster as their preferred presentation medium and some report meeting as many delegates as would attend a traditional conference session. In this year's publication we have included two papers from the poster sessions chosen as the best poster contributions by young professionals, including the overall best by G. Marra of TV Globo, Brazil.

By early May our successful authors have turned their synopses into full conference papers or posters and once again the specialists of the Technical Papers Committee carefully review every one, in most cases conveying comments back to their authors for revisions to be made. A small number of extra papers, beyond those required for the conference sessions, is requested and these will replace the few which sadly fail to meet our reviewers' expectations at this stage.

It is in June that IBC and the IET meet to plan the content of this special publication representing the best media technology papers from IBC 2010 and the IET's *Electronics Letters*. This involves yet more deliberation as, together, we skim-off our best seven papers – the crème de la crème of this year's conference contributions. We also have the important responsibility of choosing from among these the winner of the Best Paper award. This year's selected papers are:

Does size matter? The impact of screen size on stereoscopic 3DTV (Winner of the Best Paper award)

In the fast-moving area of 3D technology it is rare to find a paper that discusses new solutions to a fundamental problem, but the work by the NDS team does just that. Using clear diagrams, they describe how the depth perceived in stereoscopic displays distorts nonlinearly with screen size. For viewers with small displays the experience of stereo could prove to be a real disappointment!

The paper then examines how processing in the set-top box could compensate for this distortion and it gives two distinct solutions. First a low-cost, approximate approach and secondly, a full-blown pixel-based method requiring real-time computation of a depth map. Currently, the latter is impractical in the home but the authors foresee future stereoscopic compression standards which will also use depth maps, making more practical the implementation of their advanced correction method.

Integral 3D television using full resolution super hi-vision

Just when you thought you had caught up with the outstanding developments by NHK and its collaborators, they unveil something new. This time, with JVC Kenwood, it's 3D-UHDTV and true to their ambitious nature, this is not mere stereo, but real 3D based on a lens array method known as integral photography. The method is very well described and details of the prototype, with its 100 000 lens array, are helpfully illustrated with many diagrams and photographs. The result is a 3D image with a maximum spatial frequency of 11.3 cycles/degree that changes smoothly across a viewing angle of 24°.

Newly developed UHDTV camera system

NHK's programme of ultra-HDTV developments is truly breathtaking and this paper is the latest in an annual series of IBC papers which have successively redefined the bounds of television imaging. Here we see the latest generation of their 33 megapixel camera, looking more like a miniaturised product that a laboratory prototype. The paper describes the new four-sensor imaging, the output mapping into 16 HD-SDI streams, the camera head and CCU (a mere 3 U in height), and the sophisticated video processor. Also new is the video RAM recorder which employs a bank of 16 AVC-Intra compressors together with compression for 24 audio channels – it can hold a remarkable 2 hours of UHDTV content!

The step into the light – the EBU loudness recommendation R128

The biggest problem in television audio – finding a means of characterising loudness so that the advertisements can be held to the same perceived level as the accompanying programme – may at last have been solved. This fascinating paper by the Austrian Broadcasting Corporation examines the legacy of existing ITU standards and comes right up to date by describing the radical new recommendation soon to published by the EBU. It proposes abandonment of the established quasi peak program monitoring method and adoption of its new loudness metering and normalisation.

Wireless and fibre-optic live contribution link for uncompressed super hi-vision signals

This is another strikingly ambitious contribution from NHK's UHDTV development programme. Here they describe both a short-range wireless and long-haul fibre link capable of carrying live uncompressed UHDTV at 24 Gbit/s. The wireless link carries the 16 HD-SDI source signals in the 120 GHz band and trial results are given over a distance of 1 1/4 km. For the fibre link, an 'easy' solution would have been to employ their own transmission scheme on dark fibre, but the NHK team describes instead how it chose to convert the source data into OTU3 frames for carriage on existing leased links. Full performance curves reveal the success of this remarkable achievement.

Interactive visualisation of live events using 3D models and video textures

IBC has been fortunate to hear and publish some exceptional work from this BBC team over many years. In their latest paper they describe how conventional 2DTV may be enhanced through its combination with 3D models to provide a richer and more interactive viewing experience. Typically this might include a full 3D stadium model surrounding a live sporting event. Of course, this involves 3D graphics processing in the home but the team shows how scenes can be constructed using an existing web browser. They also describe their associated tools for precisely tracking objects and camera moves.

Robust and low-complexity detection technique for DVB-T/H receivers in fast fading channels

Pushing the limits of flexibility and spectrum efficiency with terrestrial and hand-held DVB means operating its OFDM with 8K carriers and 64 QAM modulation. This makes it very prone to inter-carrier interference which limits its receivability, especially in moving vehicles. In this extremely informative paper from CRC, Canada we learn how they have developed (excepting details currently being patented) a new method called decision-directed channel estimation. Simulated performance curves promise to allow the reception of rate half-coded signals at 180 km/h using a single antenna.

Dr Nicolas Lodge
Chairman IBC Technical Papers Committee

Selected content from IBC 2010. Winner of the Best Paper award

ISSN 2041-5923

Does size matter? The impact of screen size on stereoscopic 3DTV

K. Murray L. Chauvier S. Parnall R. Taylor J. Walker

NDS UK, Chandler's ford, Four Stoneycroft Rise, Hampshire SO53 3YU, United Kingdom
E-mail: kamurray@nds.com

Abstract: Increase the size of your television screen and the picture gets bigger - but if you have a 3D screen, should the picture also get deeper? From a single common signal, what is the effect of larger or smaller screen sizes on our 3D perception? This paper looks at what happens to the stereoscopic 3D effect as screen sizes scale. It does so based on an analysis of the geometry of stereoscopic 3DTV and argues that there is an impact as screen sizes scale. The paper looks at the theoretical potential for compensation for different screen sizes under typical viewing distances, and the challenges in performing this in low-cost set-top boxes and TV sets where the input is a stereoscopic image pair.

1 Introduction

Stereoscopic 3DTV (S3DTV) continues to maintain a high level of interest and enthusiasm from consumers, content creators and now broadcasters [1–3]. We expect to see S3DTV content for the home come from a wide range of sources. Several broadcasters have made announcements about S3DTV trials and services and, by the end of 2010, many will be transmitting on-air S3DTV content [4–6]. Then there is packaged S3DTV content using extensions to the BluRay specification and the new 3D BluRay players to carry movies. Finally yet another reason for consumers to acquire S3DTV sets are the various updates that are either already released or expected for various games platforms, together with new games that will provide a thrilling 3D gaming experience.

Matching this wide range of sources is the wide range of S3DTV sets, primarily based on shutter technologies, many of which are already on the market place [7–10]. These televisions sets currently cover a relatively limited range of sizes between 40 and 55 inch, the premium segment, but this size range will increase in the future, especially as the announced S3D capable projectors that are easily able to support 80 inch images, or even larger, become available.

This paper explores the potential consequences of this wider range of sizes for televisions on the experience that the consumer will have when watching S3DTV content. Specifically, it asks the questions if and how the S3DTV experience is affected by changing the size of a television set. These questions are posed and answered examining how the perception of the depth of an object is altered as the size of a display changes. On the basis of this, it is argued that different display sizes will provide a perceivably different experience of the same S3DTV content.

It is clearly not realistic to expect broadcasters or content packagers to produce different versions of the content for each display size, so we can assume that all sets will be receiving exactly the same broadcast or packaged signal. Games content, at least where created within the games machine and not using pre-calculated sequences, has the potential to be created specifically for the display size to which the games machine is connected. Thus, at least in theory, 3D games can automatically adapt to the size of the display by slightly altering the parameters they use to render each frame. Therefore this paper explores two potential main methods to provide corrections that could be made to the stereoscopic images to compensate for variations in display size. Each approach takes a different view on what the 'correct' behaviour should be and, as such, each has its own set of benefits and drawbacks, and these are discussed.

2 Geometry of stereoscopic 3DTV

S3DTV provides the illusion of a 3D experience by providing two views, one for each eye. Whilst there are numerous technologies that provide the separation of the images to the eye [11] (shutters, passive polarisers, parallax barriers and lenticular lens to name the four best known technologies), all rely on the way that the two separate images are interpreted by the human visual system. The basic geometry of this interpretation, and how the depth of an object is perceived, is well known [12, 13] and shown in ray diagram form in Fig. 1.

From this, we can see that the perceived depth of an object is determined by the following formula, using the symbols from Fig. 1:

$$D = \frac{e * t}{e - s}$$

There are numerous factors that affect the perceived depth of an object, including angle of view. For the remainder of the paper we shall ignore variation in the following factors:

- *t* the distance to the television is clearly a variable, but as with common broadcasting practice [14], we will assume that the viewing distance is fixed at three times the height of the display (3H). It is very rare for this to alter during viewing.

- *e* the eye separation for any given viewer is fixed, though every different person will have subtly different values for *e* [15].

- Angle of view [16]. For simplicity, it is assumed that the viewer is viewing objects as shown in Fig. 1, perpendicular to the screen. Clearly the angle will affect the view, but the angle will not change noticeably during viewing and so the impact of off-centre viewing is consistent.

2.1 Implications

The formula shows that the perceived depth is inversely proportional to the separation of the renditions. This in turn means that scaling the separation does not have a linear impact on the depth at which the object is perceived. For example, a set of objects at 4 t, 3 t and 2 t in an original screen will be perceived on a new screen of half the size at 1.6 t_{new}, 1.5t_{new} and 1.3 t_{new} where t_{new} is the new distance to the new half sized screen. The following Sections and diagrams explore the implications of this depth distortion in more detail.

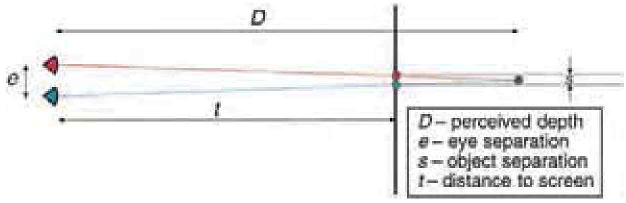


Figure 1 *S3DTV geometry diagram*

3 Depth budget and depth range

The depth budget refers to the maximum range of depths in use at any given time, the distance between the nearest and furthest objects that the viewer perceives. The choice of depth budget may be artistic, but it has a significant impact on the comfort of viewing S3DTV [17]. As a gross simplification, the use of excessive depth budget or placing objects at the limits of the depth range (especially very close to the viewer and with typical TV viewing conditions) decreases the comfort of viewing content. Thus, in general, content uses a restrained depth budget [18].

Fig. 2 shows the perceived depth placement for a range of objects based on the separation of the left and right rendition of the objects. Each object is also labelled with the separation of its left and right renditions on the screen as a multiple of eye separation. In the rest of this paper, we will show how this depth varies, or can be varied.

4 Effect of display size changes

Let us assume that we have a stereoscopic picture that provides for objects placed as shown in Fig. 2. Now let us consider taking this image and displaying it on two new screens of different sizes. Clearly, as the screen size scales, the separation scales and so the perceived depth scales. Fig. 3 shows the impact on the depth range when displayed on a screen half of the size, showing the objects at their newly labelled positions. A result of this scaling is that objects that were previously at infinity are moved to a depth of 2 t, and this is shown as ∞ in Fig. 3. In comparison Fig. 4 shows the impact on a screen twice the original intended size.

As may be seen from these two Figures, shrinking the screen compresses the depth range whereas enlarging the

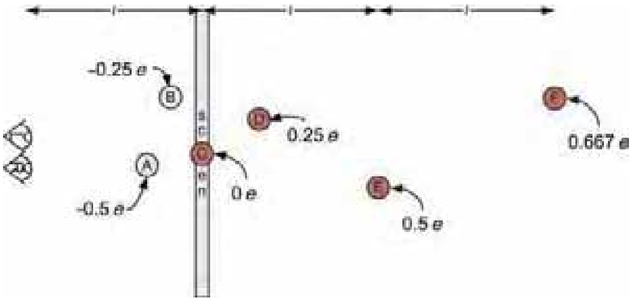


Figure 2 *Depth placement of objects and the related left–right separation*

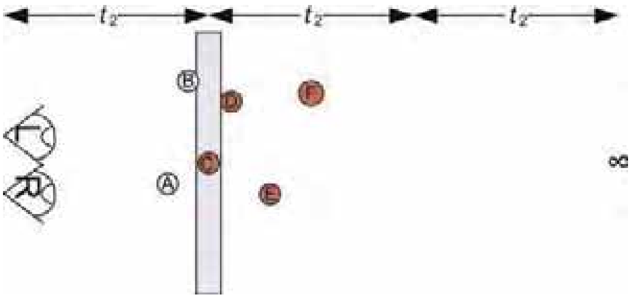


Figure 3 *Scene from Fig. 2 as perceived on a screen of half size*

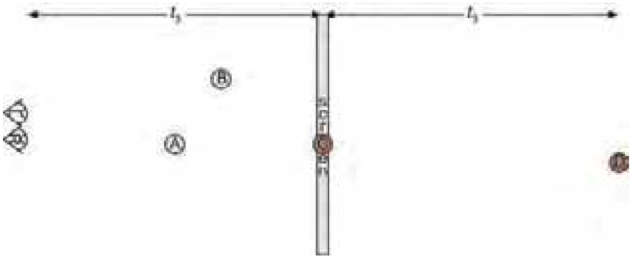


Figure 4 *Scene from Fig. 2 as perceived on a screen of double size*

screen expands the depth, but far more dramatically than just altering *t*. The perceived depths from the figures are shown in Table 1, measured both in absolute terms compared to the original display size and in relative terms.

There are several implications from scaling the images as shown in Figs. 3 and 4. First, consider an object that is moving towards the viewer or camera at a linear speed in the original image. When this object is viewed on a different sized display, the speed will no longer be perceived as linear towards the viewer, it will either be speeding up or slowing down as it approaches the viewer depending on the scaling and positioning of the object. In a very similar fashion, this will result in minor apparent size distortions.

Next, consider looking at an object. Cues, such as binocular disparity, allow the object to be accurately placed in depth by the human visual system. This depth placement is then also used by the human visual system to estimate the size of the object. We are already familiar with this effect, which is becoming known as the 'subbuteo effect' in reference to the popular old football table game, describing how football players can appear as miniature people. This effect will be strengthened by shrinking the screen size as we now have the people appearing closer to the viewer. That is, by appearing closer the human visual system will interpret the football players as being even smaller.

Finally, if extremes of depth are used in the content, scaling the display introduces another set of problems. Where the display is larger than the original, then objects can appear 'beyond infinity' (although for short term viewing, the human visual system seems to be able to accommodate this). For increased image sizes, objects close to the viewer are less of a problem since the increased (typical) viewing distance provides a degree of compensation, as shown in the table above. By contrast, where the display is smaller than the original, certain depth locations are no longer achievable. Most obviously, it is not possible to place an object at mathematical infinity. Unlike enlarging a screen, close objects can now pose a problem as they are closer to the viewer, even though the reduced separation also places them closer to the screen.

5 Depth perception correction

In the following Sections, we outline two approaches to correcting the depth impact that we introduced above: relative and absolute.

5.1 Relative depth correction

The goal for relative depth correction is to alter the stereoscopic image pair so that depths scale consistently. An example of this is shown in Fig. 5, where the original screen and objects are shown dashed, and the new scaled screen is shown solid (in this case the screen is half the size). The depths that would be seen without this correction, as

Table 1 Variation in perceived depth as screen size scales

Object		Depths in original units (t)			Display size relative depths (t_2 and t_3)	
		original	half size	double size	half size	double size
A	$e/2$	$2/3\ t$	$2/5\ t$	t	4/5	1/2
B	$e/4$	$4/5\ t$	$4/9\ t$	$4/3\ t$	7/8	2/3
C	0	t	$t/2$	$2\ t$	t_2	t_3
D	$+e/4$	$4/3\ t$	$4/7\ t$	$4\ t$	8/7	2
E	$+e/2$	$2\ t$	$2/3\ t$	∞	4/3	∞
F	$+2\ e/3$	$3\ t$	$3/2\ t$	$>\infty$	3/2	$>\infty$
∞	e	∞	$2\ t$	$>\infty$	4	$>\infty$

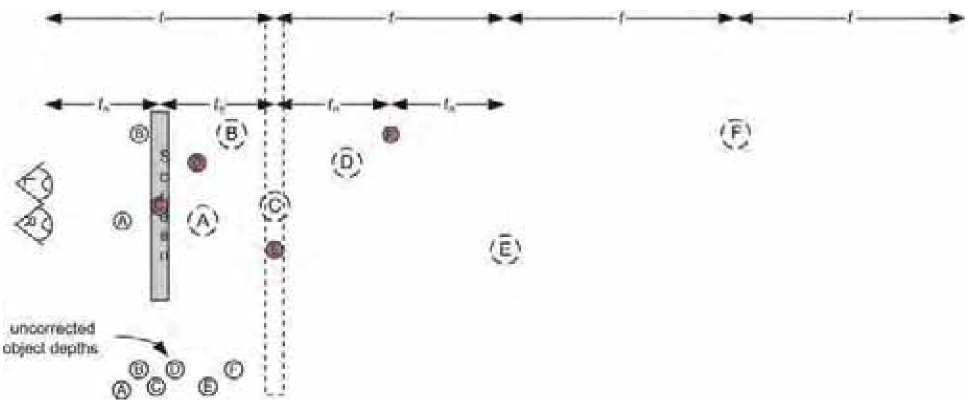


Figure 5 *Relative depth corrected example*

illustrated in Fig. 3 above, are shown in the bottom left of the diagram. This shows the desired relative depth corrections that are required to overcome the nonlinear depth compression and maintain the correctly scaled relative depths.

Given the depth equation, the desire is to retain the same absolute object separation, *s*, on the objects on the new display, even once the display is scaled. However, for a display that is scaled relative to the original target display by a factor of *f*, the separation is also scaled, so that s_{new} $s * f$. Hence, the correction we need to apply is

$$shift = s - s_{new}$$

which is

$$shift = s * (1 - f)$$

This means that the shift to be applied to an image varies across the image in relationship to the depth that is perceived for that part of the image. Therefore no single correction factor can be applied that is consistently correct.

5.2 Absolute depth correction (window on the world)

Absolute depth correction is where the stereoscopic image is adapted so that the objects appear at the same absolute distance from the viewer regardless of the change in the size of the display. In this, objects can be moved from behind to in-front of the screen, or vice versa, depending on the size changes. This is shown in Fig. 6, where the original screen is shown dashed together with a new screen that is half the size of the original, and consequently half the distance from the viewer.

If we measure the separation as a fraction of the eye separation, i.e. s $i * e$, then depth equation shown earlier becomes:

$$d = \frac{t}{1 - i}$$

The inverse, or the separation *i* required to place an object at depth *d* on a display at distance t, is:

$$i = 1 - \frac{t}{d}$$

For a screen that is scaled by a factor *f*, the new value for the distance to the screen, t_n, is t_n $t * f$. For this scaled screen using the equations above, to represent an object at depth d requires a separation $i_{corrected}$

$$i_{corrected} = 1 - (1 - i) * \frac{t_n}{t}$$

which when t_n is replaced, simplifies to

$$i_{corrected} = 1 - (1 - i) * f$$

When the image is scaled, the separation is also scaled, so i_n $i * f$, so the shift required is

$$shift = e * i_{corrected} - e * i_n$$

which, when substituted with the above becomes

$$shift = e * (1 - (1 - i) * f - i * f)$$

and which in turn simplifies to

$$shift = e * (1 - f)$$

Thus a single shift across the entire image repositions all the objects back to their original depth placements as if they were viewed on the original screen at the original depth.

5.3 Comparison of solutions

Absolute depth correction is clearly easier to apply, but it is important to consider the impact that this may have on the viewing experience and the acceptability of this. For smaller displays, it means that the content will always be distant and will occur further from screen depth. It is generally accepted that the closer to screen depth, the easier content is to view owing to lower focus-convergence disparity. The absolute depth correction solution may reduce the amount of content

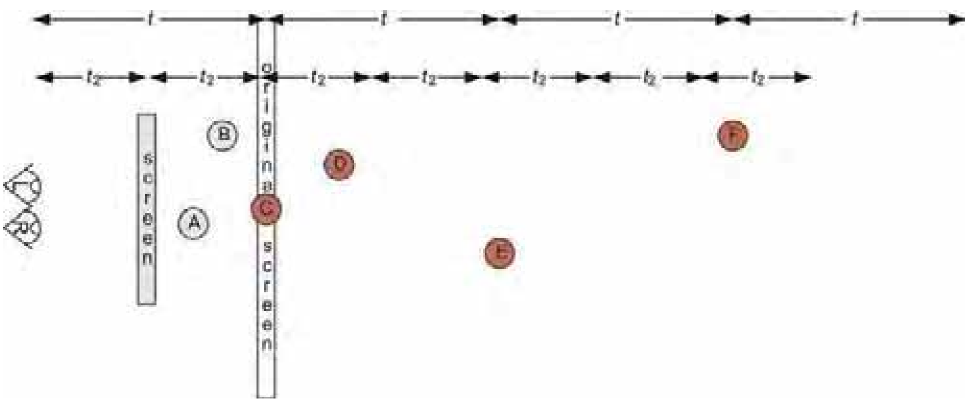


Figure 6 *Absolute depth*

that sits in this comfort zone. There is also a potential risk that making the action occur 'further' from the screen may not be as acceptable or engaging to the viewer, somewhat like the differences between watching a play from a distance, rather than being seated closer to the action. Finally, the shift required by absolute correction will result in the loss of a small amount from the edges of the picture.

Although relative correction is a more complex solution requiring computationally intensive depth map calculation [19], the authors believe that this is an important alternative and so the next Sections discuss some practical means of implementing an approximation of the relative depth correction. Clearly, however, comparative evaluations of these different solutions are required.

6 Implementing approximate relative depth correction

In the previous Sections we have seen how each separate object depth requires a different shift to be applied to compensate for the scaling of the display. Thus, an idealised solution requires an accurate depth map of the stereoscopic image and a processing stage that provides the required manipulation on a pixel by pixel basis. In theory, since an item of content could cover the entire range of depths, the shift required could vary from $e^*(1 \quad f)$ (to correct for objects at infinity) through 0 (for objects at screen depth) through to negative values of perhaps $\quad e^*(1 \quad f)$ or beyond. There is no theoretical limit to the separation that can be used for objects appearing in front of the screen. However, the convergence-focusing disparity makes perceiving objects far in front of the screen difficult. As such, except for very short durations, it is unlikely that a large negative separation will be used.

When an item of content is created, as we discussed earlier, it is rare for it to use the entire potential depth range. For any given item of content, the range of shifts to be applied will be defined by the range of depths that are represented within the stereoscopic content. Although there is still a learning experience underway to decide the most acceptable depth range to use, the current experience tends towards using a depth range from slightly behind the screen backwards, with a limit to the maximum depth. Translating this into fractions of eye separation, this currently appears to be between 0.6 *e* and about 0 or 0.1 *e* for most content (i.e. it is behind the screen in a relatively confined depth range), though clearly, this will be exceeded in certain dramatic conditions.

Fig. 7 illustrates the impact of depth over this range, shown as fractions of eye separation, *e*, on the *x*-axis, and the perceived depths over this range, in units of the distance to the display, on the *y*-axis. The solid line shows the depth range for the original display. The dashed line shows the depth range for a screen of half the size, as measured in the units of the reduced distance to the display that is a result of the smaller screen. The final dotted line shows the depths with an arbitrary constant shift of 0.15 *e*.

From this we can see that a shift can be used to minimise the overall total depth discrepancy that is introduced when a screen is scaled. The shift can be calculated in several ways, each with different properties. The simplest is to base this on the shift required to correct the mid-point of the shifts. In the case of the graph, the midpoint is 0.35, and would generate a shift of 0.175 *e*. However, the midpoint of the shifts is not the same as the midpoint of the depths, which is at a depth of 1.8 *t*, or a shift of about 0.44. This generates a correction of about 0.22 *e*.

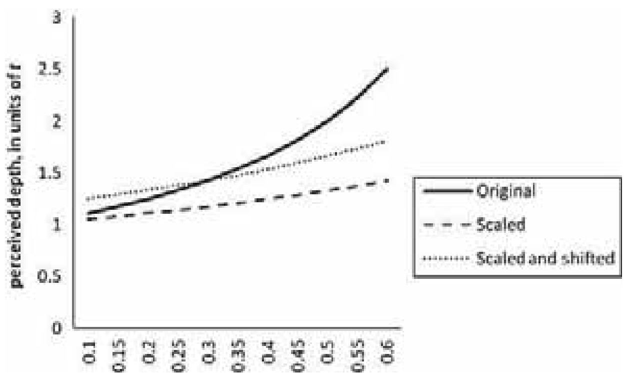


Figure 7 *Graph showing impact of constant shift on perceived depth ranges*

More complicated measures can take into account the use of each given perceived depth, i.e. it can be based on a statistical processing of the depth information. To perform this, however, one clearly needs a detailed depth map. Creating such a map is a well-known problem, with solutions that operate at real time, but the current cost of such depth (or disparity) map creation is not realistic for set-top box (STB) hardware. More realistic is the generation, and potential processing, of such a map as part of the broadcasting process. Indeed, such processing can assist several other operations in the STB such as subtitle placement by providing details of the correct depth for objects to be placed at.

Clearly further improvements are possible, for instance by combining the depth processing with scene analysis to identify areas of changing depth and prioritising those areas, or by performing regional shifts on the image. The goal here would be to minimise the errors over the range that is seen as being of dramatic importance.

7 Concluding remarks

The success of S3DTV will be in giving every viewer a high-quality experience without eyestrain, no matter what sized sceen is being used in the home. In this paper we have explored the effect of changing screen size for a given input signal and show that this will have an unintended impact on the perceived depth range of the resultant image. It would be possible, but clearly not viable, to transmit multiple streams for specific screen sizes. We show that it is also possible to provide local screen size compensation in the home effectively and efficiently from a common input signal, and propose two alternative approaches to this technique. Clearly, further study on the acceptability of these solutions is needed, but we assert that they do provide the basis by which 3D screen size compensation may be effected at a low cost.

In the longer term, it is likely that coding techniques for S3DTV and multiview 3DTV will be based on, or utilise, depth maps or something similar. This will then open up the possibility of a full and accurate relative depth correction solution.

8 References

[1] BBC: 'Broadcasting first with 3D Rugby, BBC News Website', 10 March 2008, http://bit.ly/crCGsy

[2] RICHARD SANDOMIR: 'In 3-D, Masters does have extra dimension', *The New York Times*, 31 March 2010, http://nyti.ms/bjuru5

[3] FIFA Press Release: '2010 FIFA World Cup to pioneer 3D technology', 3 December 2009, http://bit.ly/bbOK2k

[4] ESPN Press Release: 'ESPN 3D to show soccer, football, more', 5 January 2010, http://es.pn/8zeFwG

[5] BSkyB Press Release: 'Sky 3D to Launch with Manchester United vs Chelsea on Saturday 3rd April, 19' March 2010, http://bit.ly/djRmje

[6] DirectTV Press Release: 'DirectTV is the First TV Provider to Launch 3D', 6 January, 2010, http://bit.ly/7xKbhA

[7] Samsung Press Release: 'Samsung announces true 3D capability for multiple screen technologies', 26 January 2010, http://bit.ly/bdNHNN

[8] Sony Press Release: 'Complete 3D experience hits UK shops – Sony announces lens to living room offering', 11 June 2010, http://bit.ly/aF2FLh

[9] LG Press Release: 'LG brings the first passive 3d-Ready TV to UK consumers', 31 March 2001, http://bit.ly/dq51RM

[10] Panasonic Press Kit: 'Panasonic 3D Full HD system launch', March 2010, http://bit.ly/dcDACk

[11] Mark Schubin: 'What is 3D and why it matters'. NCTA 2010 Spring Technical Forum, May 2010

[12] MATT DEJOHN, WILL DREES, DAVID SEIGLE, JIM SUSINNO: 'Stereoscopic geometry of 3D presentations', *In-three*, http://bit.ly/d4ENXq

[13] DAVID WOOD: 'The truth about stereoscopic television'. Proc. 2009 Int. Broadcasting Convention, September 2009

[14] ITU-R Recommendation BT.500-11: 'Methodology for the subjective assessment of the quality of television pictures', 2002

[15] NEIL DODGSON: 'Variation and extrema of human interpupillary distance', *Proc. SPIE – Stereoscopic Displays and Virtual Reality. Sys.*, 2004, **5291**, http://bit.ly/brHuaH

[16] Martin Banks: 'Stereoscopic vision: how do we see in 3D?', NAB 2009, http://bit.ly/bcX2DN

[17] DAVID HOFFMAN, AHNA GIRSHICK, KURT AKELEY, MARTIN BANKS: 'Vergence-Accomodation Conflicts Hinder Visual Performance and Cause Visual Fatigue', *J Vis.*, 2008, **8**, (3), pp. 1–30, 33, http://bit.ly/aNbkc0

[18] BERNARD MENDIBURU: '3D movie making: stereoscopic digital cinema from script to screen', *Focal Press*, 2009

[19] DANIEL SCHARSTEIN, RICHARD SZELISKI: 'A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms', *Int. J. Comput. Vis.*, 2002, **27**, (1–3), pp. 7–42, http://bit.ly/clRvWQ

Selected content from IBC 2010

ISSN 2041-5923

Integral 3D television using a full resolution super hi-vision

J. Arai[1] *M. Kawakita*[1] *H. Sasaki*[1] *H. Hiura*[1] *M. Miura*[1] *M. Okui*[1] *F. Okano*[1] *Y. Haino*[2] *M. Yoshimura*[2] *M. Furuya*[2] *M. Sato*[2]

[1]*NHK (Japan Broadcasting Corporation), 1-10-11 Kinuta Setagaya-ku, Tokyo 157-8510, Japan*
[2]*JVC KENWOOD Holdings, Inc., 58-7, Shinmei-cho, Yokosuka, Kanagawa 2398550, Japan*
E-mail: arai.j-gy@nhk.or.jp

Abstract: An integral 3D television using a full resolution super hi-vision system is presented. The system uses a device having 7680 pixels in the horizontal direction and 4320 pixels in the vertical direction for each of the red, green, and blue channels. A lens array is configured with 400 lenses in the horizontal direction and 250 lenses in the vertical direction. The system is designed to ensure a maximum spatial frequency of 11.34 cycles/degree in the horizontal direction when the display is observed from three times the display height. The authors have confirmed the generation of a 3D image with an appearance that varies in a natural manner according to the position of the viewer without using 3D glasses.

1 Introduction

Since early times, humans have been trying to visually represent a space having depth. It has been pointed out that the Lascaux cave paintings show that a perspective drawing method that could represent a 3D space was used as early as 15 000 years ago. The 3D image technique that uses 3D glasses and has been popular in recent years was first attempted in the 1600s [1]. Integral photography (IP), which is the basic principle of integral 3D television, was proposed in 1908 [2]. The fact that such attempts have continued indicates that people want a faithful representation of the space in which they live.

Broadcasting, on the other hand, can be said to contribute to the forming of a healthy society by a wide range of offerings in information, entertainment and education. The offering of natural 3D video as a broadcasting service, has the potential to create a new culture. Anticipating that such a service would enrich the lives of viewers, we have taken on the challenge of developing 3D television.

We consider the three points described below to be the basic policy for 3D television. The first point is enjoyment of 3D video without having to use special 3D glasses; the second is enjoyment of 3D video from any posture; and the third is implementation of the process from shooting to display in real-time. Considering that 3D video broadcasting would be enjoyed at many different places, at home and outdoors, we want a system that allows viewers to watch the programme from any posture without wearing 3D glasses. Also, a system that allows live presentation is essential for a broadcasting service. Integral 3D television satisfies these three basic policy items. In integral 3D television, three-dimensional information is represented in two dimensions for both shooting and display. Accordingly, to reproduce the image with good quality requires high resolution for both shooting and display devices.

A super hi-vision (SHV) system can represent motion pictures at the highest resolution at the current time [3–5]. We therefore introduce here integral 3D television that uses full resolution SHV technology. In Section 2, we briefly describe the 3D display techniques that have been proposed so far. In Section 3, we explain the principle and display characteristics of integral 3D television. In Section 4, we introduce an experimental integral 3D

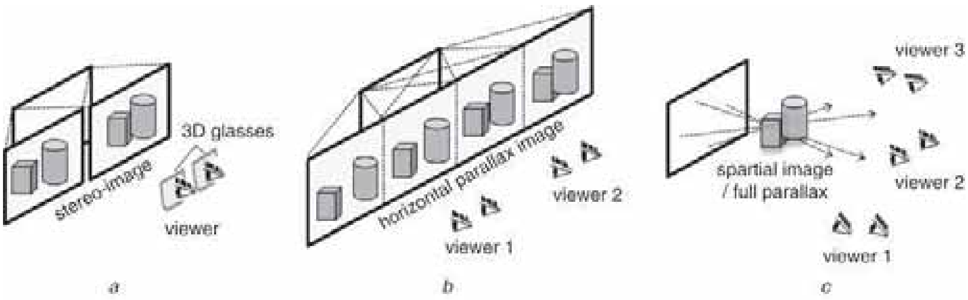


Figure 1 *3D effect*
a Stereogram
b Parallax panoramagram
c Integral photography (IP)

television that uses a full resolution SHV. The final section concludes this paper.

2 3D display technology

The difference in a scene as seen by the left and right eyes is a major factor in producing the 3D effect in vision. That difference is called binocular parallax. Accordingly, 3D display requires at least two images, one for the left eye and one for the right eye. Furthermore, these two images must be accurately separated and presented to the left and right eyes.

Methods that use 3D glasses, separate the left and right images by using polarised glasses or the optical shutter (Fig. 1*a*). Since that approach is technically simple and produces good image quality, it is widely used in movie theatres and other such venues. Nevertheless, only one pair of left and right images is displayed, so even if the viewer moves, the 3D video that is seen does not change. Another factor in our sensing of the 3D effect is motion parallax. Motion parallax is a difference in the scene that is being viewed created by movement of the viewpoint. Generally, methods that use 3D glasses cannot provide motion parallax.

One 3D display that can present motion parallax is the parallax panoramagram (Fig. 1*b*). In this method, the subject is photographed from multiple positions in the horizontal direction (four, for example) rather than using a single left–right image pair. When the image is displayed, a slit array or lenticular lens controls the directions from which the multiple images can be seen. Nevertheless, the different images can only be presented in the horizontal direction, so viewers must keep their two eyes level to obtain the 3D effect. This constraint is the same for 3D glasses.

Known ways to achieve 3D display without using 3D glasses and without restricting the posture of viewers are Integral photography (IP) (Fig. 1*c*) and holography. IP in particular offers the advantage of using ordinary natural light to photograph the subject and display the 3D image. As a result of these advantages, we are developing integral 3D television on the basis of integral photography.

3 Integral 3D television

3.1 Principles

The French physicist Lippmann proposed IP as a 3D photographic technique. Integral 3D television is based on IP and achieves subject shooting and image display in real time. IP uses a lens array composed of many convex lenses and film to capture the images (Fig. 2). In the capturing stage, as shown in Fig. 2*a*, each individual convex lens creates an image of the subject, so the number of images captured by the film is the same as the number of lenses. We call the convex lenses ‘elemental lenses’ and the images of the subject produced by the elemental lenses ‘elemental images’. In displaying the image, a lens array is placed in front of the film on which the elemental images are recorded as shown in Fig. 2*b*. The light rays from each elemental image then pass through an elemental lens and return in the opposite direction of the incoming light rays during capturing stage, so the light emanating from the subject is reproduced. The result is that the viewer can see 3D image without having to wear 3D glasses. Ordinary natural light is used both when capturing the subject and when displaying the 3D image.

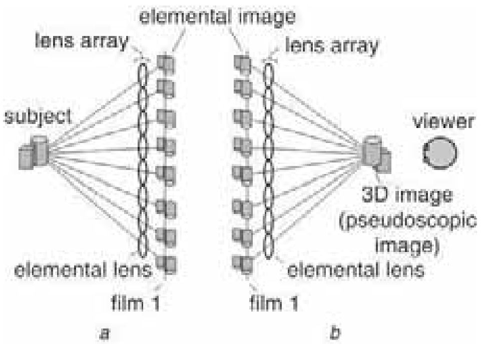


Figure 2 *Principles of integral photography*
a Capturing stage
b Displaying stage

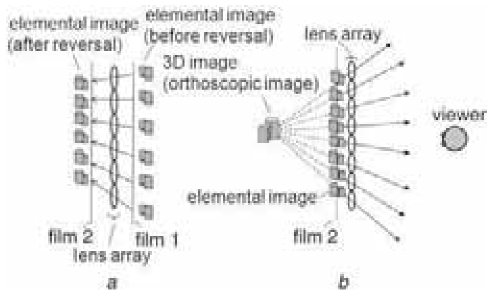

elemental image (after reversal)
elemental image (before reversal)
lens array
3D image (orthoscopic image)
viewer
elemental image
film 2
film 1
lens array
film 2
a
b

Figure 3 *Display of an orthoscopic image*

a Reversal of elemental images
b Displaying 3D images

With the configuration shown in Fig. 2, the depth of the reproduced 3D image is reversed relative to the subject, creating what is called a pseudoscopic image. For example, in Fig. 2*a*, the cylinder is in front of the block as seen from the film, but in Fig. 2*b* the cylinder appears to be further away from the viewer than is the block. The pseudoscopic image effect can be prevented by reversing the point symmetry of each elemental image. This reversal process can be accomplished by two-step capturing. First, as shown in Fig. 2*a*, the elemental image is recorded on film 1. Next, the elemental image is transferred to film 2 via a lens array as shown in Fig. 3*a*. When a lens array is placed in front of the film from the second capturing stage as shown in Fig. 3*b*, a 3D image with correct depth display is produced.

The problems of achieving real-time capturing and display in IP arise from the use of film and the two-step capturing required to eliminate pseudoscopic images. The first problem can be averted by replacing the film with a CCD, or other such electronic capturing device, and using an LCD panel or other such display device. The second problem can be averted by using a gradient-index lens to generate an elemental image that is equivalent to the elemental image produced by two-step capturing, as illustrated in Fig. 4. The refractive index of a gradient-index lens varies parabolically in the radial direction from the centre of the lens.

Thus, integral 3D television can be realised by replacing film with electronic imaging devices and using a gradient-index lens for the elemental lens used for capturing [6].

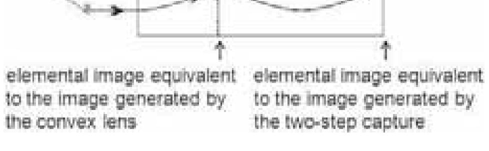

gradient-index lens
subject
elemental image equivalent to the image generated by the convex lens
elemental image equivalent to the image generated by the two-step capture

Figure 4 *Image formation by a gradient-index lens*

3.2 Resolution of displayed 3D image

A 3D image generated by integral 3D television can be thought of as a stack of flat images that are superimposed in the depth-wise direction. Imagine an elemental image projected by an elemental lens onto a position where a 3D image is reconstructed, as shown in Fig. 5. In this case, assume that the maximum frequency of stripes per radian of the projected image is α, which is called the 'projection spatial frequency'. When a 3D image is viewed from a distance L from the lens array, as shown in Fig. 6, the maximum frequency of stripes per radian is β, which is called the 'viewing spatial frequency'.

With a 3D image, therefore, the projection spatial frequency α is affected by aberrations owing to focusing errors of the elemental lenses, in addition to pixel pitch and the diffraction limits of the elemental lenses. If the pixel pitch of the display device is p, the spatial frequency α_p of the elemental image projected by each elemental lens is given by

$$\alpha_p = |g|/2p \qquad (1)$$

where g denotes the distance from the lens array to the display device. Assume that the spatial frequency of the diffraction limit of the elemental lens is α_d and the limiting spatial frequency after consideration of aberrations of the elemental lens is α_e. The limiting spatial frequency of the elemental image projected by the elemental image in this case, in other words the projection spatial frequency α, is given by

$$\alpha = \min[\alpha_p, \alpha_d, \alpha_e] \qquad (2)$$

The elemental images projected at the determined spatial frequency α are combined to generate the 3D image.

The viewing spatial frequency β, which is the spatial frequency when the generated 3D image is viewed from the position of the observer, is also affected by the depth-wise position of the 3D image, is given by

$$\beta = \alpha(L - z)/|z| \qquad (3)$$

where L denotes the viewing distance (distance from the lens array to the viewer), z denotes the 3D image distance (distance from the lens array to the 3D image), the right

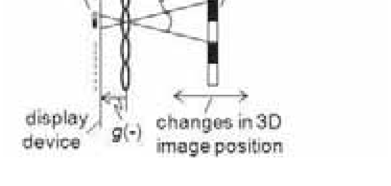

lens array
elemental image projected by elemental lens (= 3D image)
elemental image
1/α(rad.)
display device
g(-)
changes in 3D image position

Figure 5 *Projection spatial frequency*

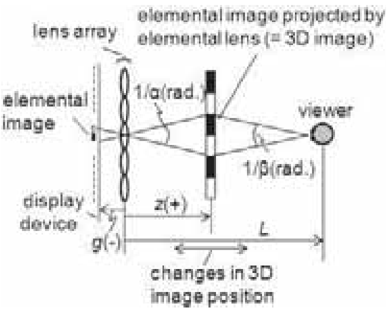

lens array
elemental image projected by elemental lens (= 3D image)
elemental image
1/α(rad.)
viewer
1/β(rad.)
display device
z(+)
g(-)
L
changes in 3D image position

Figure 6 *Viewing spatial frequency*

side from the lens array takes positive values, and the left side takes negative values, as shown in Fig. 6.

In addition, when a 3D image is displayed by an integral method, it is necessary to consider the pitch of the elemental lenses that make up the lens array. The distance from the display device that displays the elemental image by the integral method to the lens array is arranged in such a manner as to approximately coincide with the focal distance of the elemental lenses, as shown in Fig. 7. In that state, parallel light rays are output from each elemental lens, the bundle of parallel light rays from each elemental lens are superimposed, and the 3D image is generated as shown in Fig. 8. As a result, when the 3D image is viewed from the position of the viewer, the 3D image is sampled at the pitch of the elemental lenses. From consideration of the resultant geometric optics, the maximum spatial frequency of the 3D image is constrained to the Nyquist frequency determined by the pitch p_L of the elemental lenses, as expressed by

$$\beta_n = L/2p_L \tag{4}$$

where this spatial frequency β_n is called the 'maximum spatial frequency'.

From the above, the limit of the spatial frequency for a 3D image generated at any arbitrary depth is constrained to a low value in comparison with the maximum spatial frequency β_n determined by the elemental lens pitch alone and the viewing spatial frequency β calculated from (3), as expressed by [7, 8]

$$\gamma = \min[\beta_n, \beta] \tag{5}$$

where the spatial frequency γ expressed by (5) is called the 'limiting spatial frequency'.

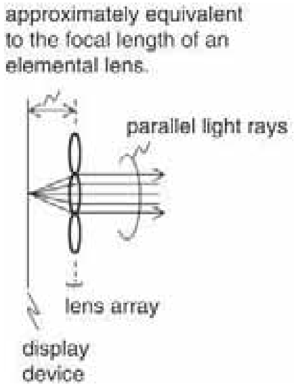

approximately equivalent to the focal length of an elemental lens.
parallel light rays
lens array
display device

Figure 7 *Output of parallel light rays*

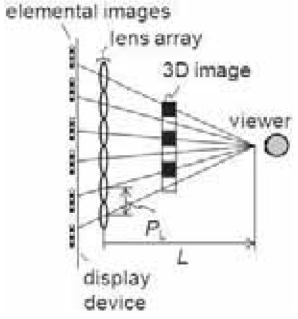


Figure 8 *Maximum resolution*

3.3 Viewing area of displayed image

With integral 3D television, a 3D image that depends on the position of the viewer is visible, whether the viewer is to the right or left or above or below. It should be noted, however, that the range (viewing area) within which the viewer can move is limited to the region within which the light from a certain elemental image is output by the corresponding elemental lens. If the shape of the elemental image is circular, by way of example, the viewing area is conical. A section through the viewing area formed by the elemental images and elemental lenses is shown in Fig. 9. In the figure, the angle Ω denotes the expanse of the viewing area and is called the 'viewing angle', which is expressed as follows

$$\Omega = 2\tan\ ^{1}(p_L/2|g|) \tag{6}$$

4 Experimental device

The configuration of an integral 3D television that uses a camera and projector designed for full resolution SHV is shown in Fig. 10 and its specifications are given in Table 1. First of all, a depth control lens is used in the image capture system to generate a real image of the subject. This enables adjustment of the depth-wise position

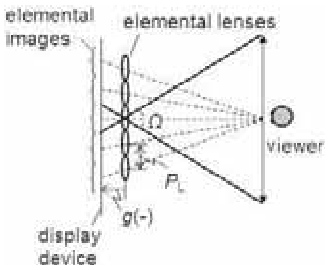


Figure 9 *Viewing area*

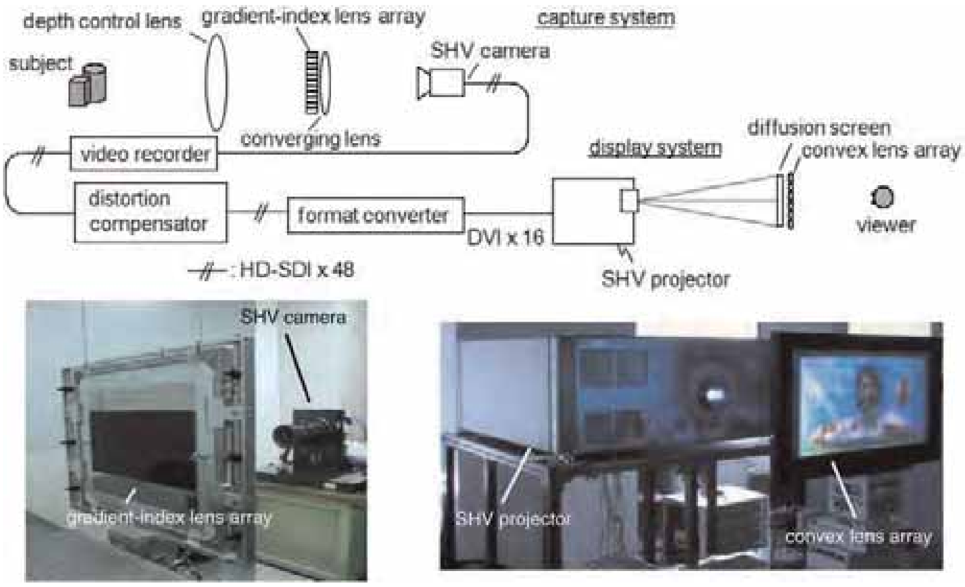


Figure 10 *Configuration of integral 3D television*

of the displayed 3D image. If, for example, the real image of the subject is generated on the camera side with respect to the lens array, the displayed 3D image is generated in front of the lens array. A group of elemental images relating to this real image is obtained by the image capture camera. To avoid the problem of pseudoscopic images, the lens array is configured by using gradient-index lenses. A converging lens is provided between the lens array and the camera, with the objective of guiding the light from the lens array efficiently into the camera. The display device uses a projection device to project the group of elemental images obtained by the camera onto a diffusion screen, and

Table 1 Specifications of integral 3D television

Capture system			Display setup		
Television camera	pixel structure pixel count (active)	CMOS 3 transistors 7680(H) × 4320(V) × 3 One each for green, blue, and red.	projector	panel pixel count (active)	1.75 in. LCOS 7680(H) × 4320(V) × 3 one each for green, blue, and red.
	frame frequency	60 frames/second		frame frequency (active)	60 frames/second
	scanning	progressive		scanning	progressive
	capture lens	focal length 61.59 mm		projection lens	focal length: approx. 59.31mm
				projection size	diagonal angle: approx. 26 in.
Lens array	type	gradient-index lens	lens array	type	convex lens
	number of lenses	400(H) × 259(V)		number of lenses	400(H) × 250(V)
	lens diameter	1.085 mm			
	lens pitch	1.14 mm (horizontal direction)		lens pitch	1.44 mm (horizontal direction)
	focal length	2.65 mm		focal length	2.745 mm
	arrangement	delta array		arrangement	delta array
Converging lens	focal length	800 mm	viewing angle		28 degrees (designed value)

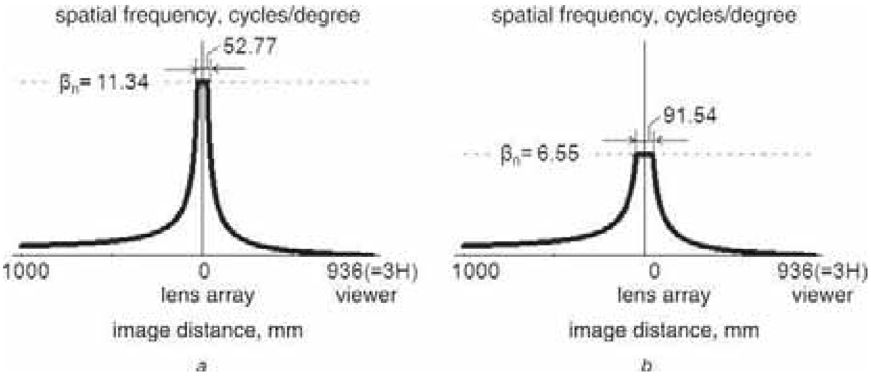


Figure 11 *Limiting spatial frequency (γ) of 3D image*

a Horizontal direction
b Vertical direction

generates a 3D image by placing a lens array configured of convex lenses on the front surface. If there is any distortion in the images projected on the diffusion screen, the generated 3D image will deteriorate [9]. With this system, deterioration of the 3D image is avoided by electrical compensation for distortion [10]. The spacing between the diffusion screen and the lens array is arranged to be approximately the same as the focal distance of the convex lenses. With SHV, a 4320-scanning-line device is used for each of the red, green, and blue channels.

The results of calculating the limiting spatial frequency γ expressed by (5) are shown in Fig. 11. The lens array for display that is shown in Table 1 is configured of elemental lenses in a delta array. Thus the elemental lens pitch differs in the vertical and horizontal directions. The maximum spatial frequency β_n was calculated from consideration of the elemental lens pitch in the horizontal direction in Fig. 11*a* and the elemental lens pitch in the vertical direction in Fig. 11*b*, and the viewing distance was set to three times the display height. With this experimental device, the viewing spatial frequency β is determined by the pixel pitch, and the pixel pitch was calculated as 75 μm in Fig. 11. This corresponds to a situation in which a 2D image of 4320 scanning lines is projected on an approximately 26-inch screen.

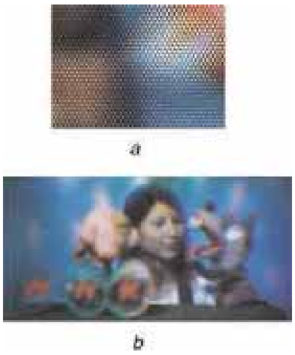


Figure 12 *Captured elemental images and displayed 3D image*

a Enlarged portion of elemental images
b Displayed 3D images

Deterioration of the resolution of a deep 3D image in this prototype device was caused by coarseness of the pixel pitch of the image. If it is assumed that a spatial frequency on the same level as that for hi-vision is required as the value of the maximum spatial frequency β_n, this would be equivalent to 30 cycles/degree, but the maximum spatial frequency β_n of this experimental device was 11.34 cycles/degree, even after considering the elemental lens pitch in the horizontal direction. Thus it is necessary to make both the elemental lens pitch and the pixel pitch finer, in order to generate 3D images of a practicable level.

We performed capture and display experiments using the experimental device. An enlarged portion of the captured group of elemental images is shown in Fig. 12*a*, and a 3D image generated by the display system is shown in Fig. 12*b*. Figs. 13*a–d* show how the appearance when the 3D image is viewed from the upper, left, right and lower positions. Fig. 14 shows the appearance when a diffuser plate was placed on a 3D image. As a reference, Fig. 14*a* shows the appearance without a diffuser plate. Figs. 14*b* and *c* show the appearance when a diffuser plate was placed

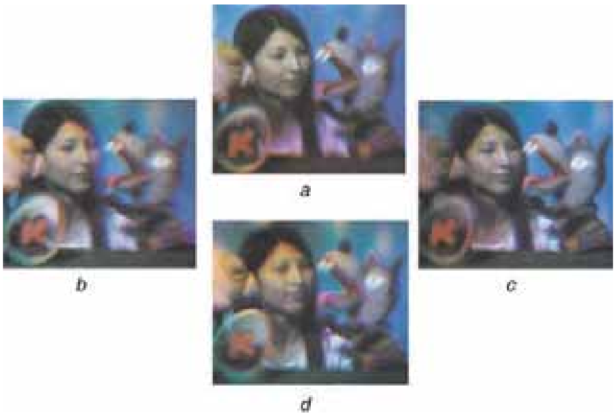


Figure 13 *3D images taken from different view positions*

a Upper view
b Left view
c Right view
d Lower view

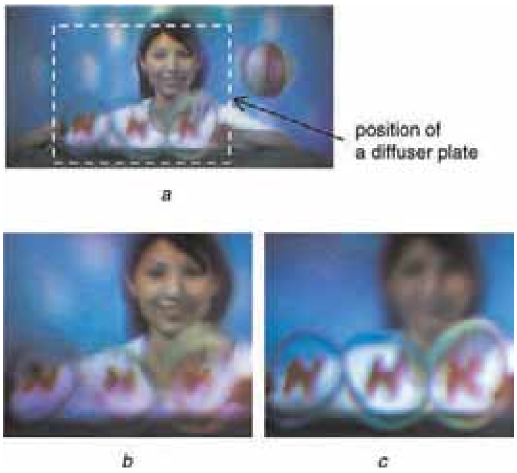


Figure 14 *Appearance when diffuser plate is placed on a 3D image*

a Appearnce without a diffuser plate
b A diffuser plate is placed over the lens array
c A diffuser plate is placed in front of the lens array

over the lens array ('human face') and on the 3D image displayed on the front surface of the lens array ('NHK'), respectively. It is clear that a 3D image is generated as a spatial image, which is generated at the same depth-wise position as the diffuser plate, which is projected clearly onto the diffuser plate. Note that the viewing angle calculated from (6) was approximately 28 degrees, and an angle of 24 degrees was measured in practice.

5 Conclusions

We have developed a new integral 3D television system that uses full resolution SHV devices that have a resolution of 7680 horizontal pixels and 4320 vertical pixels. The lens array has 400 lenses in the horizontal direction and 250 lenses in the vertical direction. The results of capturing and display experiments confirm 3D video reproduction that changes smoothly with the viewing position and 3D video that reproduces the actual spatial image. Integral 3D television promises a 3D display that allows enjoyment of natural 3D video.

The measured viewing angle of the experimental equipment is 24 degrees and the maximum spatial frequency is 11.3 cycles/degree. When viewing hi-vision from the standard viewing distance, the spatial frequency is equivalent to 30 cycles/degree. In future work, we intend to improve 3D image quality through reduction of both the elemental lens pitch and the pixel pitch.

6 Acknowledgments

The authors would like to thank Kohji Mitani, Takayuki Yamashita, Ryohei Funatsu and Yuji Nojiri of NHK for their technical support in developing the SHV camera system.

This work was supported in part by the National Institute of Information and Communications Technology (NICT).

7 References

[1] OKOSHI T.: 'Three-dimensional imaging techniques' (Academic Press, 1971)

[2] LIPPMANN M.G.: 'Épreuves réversibles donnant la sensation du relief', *J. Phys.*, 1908, **7**, pp. 821–825

[3] YAMASHITA T., HUANG S., FUNATSU R., *ET AL.*: 'Experimental color video capturing equipment with three 33-megapixel CMOS image sensors'. Proc. SPIE, January 2009, pp. 72490H1–72490H10

[4] NAGOYA T., KOZAKAI T., SUZUKI T., *ET AL.*: 'The D-ILA device for the world's highest definition (8K4K) projection systems'. Proc. IDW, December 2008, pp. 203–206

[5] ITU-R BT.1769: Parameters values for an expanded hierarchy of LSDI image formats for production and international programme exchange. ITU-R, 2006

[6] OKANO F., ARAI J., HOSHINO H., *ET AL.*: 'Three-dimensional video system based on integral photography', *Opt. Eng.*, 1999, **38**, (6), pp. 1072–1077

[7] HOSHINO H., OKANO F., ISONO H., *ET AL.*: 'Analysis of resolution limitation of integral photography', *J. Opt. Soc. Am. A.*, 1998, **15**, (8), pp. 2059–2065

[8] ARAI J., HOSHINO H., OKUI M., *ET AL.*: 'Effects of focusing on the resolution characteristics of integral photography', *J. Opt. Soc. Am. A.*, 2003, **20**, (6), pp. 996–1004

[9] KAWAKITA M., SASAKI H., ARAI J., *ET AL.*: 'Geometric analysis of spatial distortion in projection-type integral imaging', *Opt. Lett.*, 2008, **33**, (7), pp. 684–686

[10] SASAKI H., KAWAKITA M., ARAI J., *ET AL.*: 'Analysis and compensation of spatial distortion in integral three-dimensional imaging'. Proc. Third Int. Universal Communication Symp., December 2009, pp. 64–69

Selected content from IBC 2010

ISSN 2041-5923

Newly developed UHDTV camera system

K. Arai   S. Mitsuhashi   D. Ito   H. Fujinuma
R. Funatsu   T. Kikkawa

NHK (Japan Broadcasting Corporation), 1-10-11 Kinuta Setagaya-Ku, Tokyo 157-8510, Japan
E-mail: arai.k-fs@nhk.or.jp

Abstract: NHK, the Japan Broadcasting Corporation, is researching and developing an ultra-high definition television (UHDTV-2) system that has 16 times the number of pixels compared to high definition television (HDTV) as the next generation broadcasting standard. The camera, lenses, compressed flash memory recorder, and downconverter newly developed change production workflow with RAW data acquisition. Moreover, the downconverter, 8K to 4K (UHDTV-1) or 8K to 2K, makes it possible to use UHDTV contents for HDTV programmes. In this paper, the authors present the specifications of this next generation UHDTV camera system and total workflow with RAW data. In addition, the authors discuss two UHDTV lenses, with wide angles of 100 degrees and 10 times the zoom ratio.

1 Introduction

NHK (Japan Broadcasting Corporation) has been strongly promoting the standardisation and development of super hi-vision (ultra-high definition television, UHDTV [1]) technology that has 33 million pixels, 16 times as many as high definition television (HDTV), which gives the viewer the feeling of being at a live performance. We recently developed a new UHDTV system that makes full use of the latest digital and compression technology, thus permitting a variety of useful functions and ensuring high performance. Moreover, we also placed great importance on the size of the camera system so that it can operate like an ordinary HDTV camera. This paper discusses the compact UHDTV camera, the high-resolution lenses and the flash memory compressed recording system. In addition, we also introduce our UHDTV production workflow.

2 Super hi-vision (UHDTV)

Super hi-vision, with 4320 vertical pixels and 7680 horizontal pixels, has 16 times as many pixels as an ordinary HDTV system. The total number of pixels is 33 million, which means that we can show images with a higher resolution than that of a digital still camera. These large numbers of pixels make it possible to give the viewer the sense of being at a live performance.

3 Dual green super hi-vision

In this Section, we explain the features of the four sensors our UHDTV system adopts in the latest UHDTV compact camera.

First, we need to give a brief explanation about the ordinary HDTV camera. The HDTV camera used for broadcasting uses three full-specification sensors. 'Three' sensors mean that the camera has three sensors, and three also means that the sensors are for red, blue and green. Full specification means that each sensor has 2 megapixels (1920 × 1080). Therefore a camera with three full-specification sensors will convert optical signals from prism to electrons using 2 megapixel sensors (charge coupled device (CCD) or complementary metal oxide semiconductor image sensor (CMOS)).

The four-sensor UHDTV camera is the most likely system for realising the UHDTV camera system now. The advantage of this system is that we can reduce the size and weight of the camera while still keeping the UHDTV resolution. Four sensors mean that the camera has G1, G2, R and B CMOS sensors. In addition, the camera has two sensors for green channels. That is why we named this method the 'dual green' UHDTV system. Each sensor has 8 megapixels and has one-fourth of the full-specification UHDTV system. The sensor size of the dual HDTV system is 1.25 inch, making it smaller

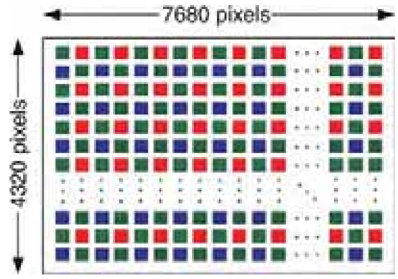


Figure 1 *Pixel array of 4CMOS SHV sensor*

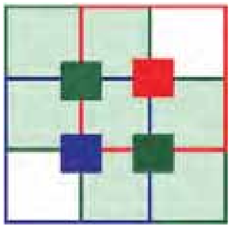


Figure 2 *4CMOS Bayer structure*

than full specification (2.5 inch) [2]. The smaller sensor can reduce the size of optical parts like the prism and lens, so this is one advantage concerning the issue of size.

Next, we will show how we can get a UHDTV signal from four 8 megapixel sensors. The pixel offset method is key technology. Fig. 1 shows the pixel offset method. Two green sensors (G1 and G2) are placed diagonally, B is placed next to G1, and R placed next to G2. With this method, the resolution of the luminance signal calculated with four sensors will be nearly equal to 33 million pixels. In addition, with a UHDTV sensor using a Bayer structure, each pixel offset is spatially only half a pixel (approximately 2 μm), as shown in Fig. 2. The reason there are two green channels is that humans are most sensitive to 540 nm and more green brings higher resolution.

The signal from this camera is mapped on 16 HD-SDI (SMPTE-292M) called dual green SDI (Fig. 3). We turn now to this dual green SDI signal. First, the signal from each CMOS sensor has 3840 × 2160 pixels, which are divided into 4 areas of a 1920 × 1080 signal. As flame rates are 60 fps, we can treat these like four colours and four areas of 1080 60 P signals. In order to map each 1080 60 P signal on HD-SDI, odd lines use the Y area, and even lines use the C area.

Finally, we should mention the full-specification UHDTV camera system that our research and development department is developing now. The full-specification UHDTV camera has three 33 megapixel CMOS sensors that detect input through prisms.

4 Super hi-vision camera

The UHDTV camera we recently developed is based on an '8 million pixel four CMOS pixel offset method'. The camera system is composed of three parts: a UHDTV camera head with a built-in optical device (Fig. 4), a camera control unit (CCU) with a built-in optical device, and a video processor (VP) correcting chromatic aberration, adding DTL (detail), applying gamma, white clip and knee. The size of the camera head has been drastically reduced compared with former ones. It has only half the volume of previous camera heads. In addition, we do not need a separated optical device.

The height of the newly developed CCU is only 3 U (19 inch rack). The former CCU had 30 memories to calculate fixed pattern noise (FPN), and reallocate the signal from the sensors in a dual green structure. The new camera needs only three huge-volume memory chips. Besides that, the calculation for FPN cancellation is done in the camera head.

The height of the newly developed VP is reduced to only 4 U and needs very low power consumption. The specifications for the new camera system are shown in Table 1.

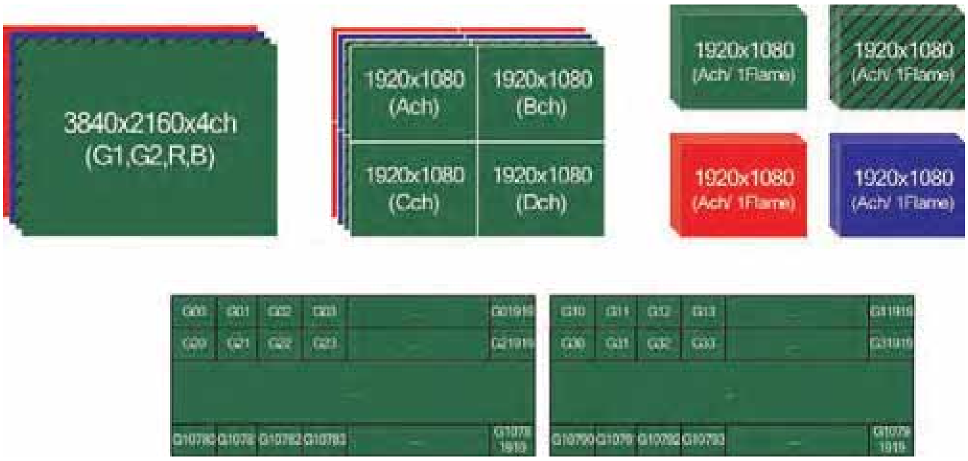


Figure 3 *Mapping of UHDTV Video on HD-SDI (SMPTE-292M)*

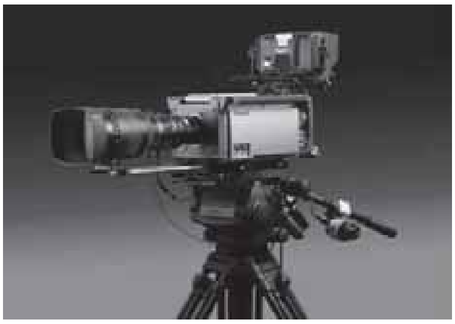


Figure 4 *UHDTV camera head*

Table 1 Specifications for UHDTV camera

Item	UHDTV camera
Sensor	8 megapixels CMOS 4 sensors (G1G2 BR pixel offset)
Pixel number	7680 × 4320
Pixel size	4.2 μm × 4.2 μm
Format	60 fps progressive
Optical	1.25 inch
Sensitivity	F4@2000lx: D 200% mode
	F5.6@2000lx: D 400% mode
S/N	57.2 dB: D 200% mode
	50.8 dB: D 400% mode
Dynamic	200%, 400% selectable
Weight	head: 19.5 kg CC: 26.3 kg
Power consumption	320W: head & CCU & VF
	500W: VP
Audio	AES/EBU input: 24ch
	analogue input: 2ch
	MADI input (HD-SDI trunk): 1ch

5 New functions of camera head

Our newly developed UHDTV camera with a very high resolution of 33 millions pixels can be operated like a HDTV camera. To operate like a HDTV camera, this camera has many new functions. The camera head has a view finder with downconverted HD resolution. The camera operator can magnify and cut out from 8 selected areas on a UHDTV signal to support forces control by using the control panel on the camera head (Fig. 5). View finder signals created by the digital signal processor (DSP) on the camera head have no delay, and it is possible to apply an all video process. The input and output signals on the camera head, focused on a variety of operational styles, are two channels of analogue audio, 24 channels of AES/EBU, and a HD-SDI video trunk. The interface between the newly developed zoom lens with 10 times the zoom ratio and a wide-angle 100 degrees lens needs only one serial command cable, including for power supply. Care has also been taken about the design, which has a round-shaped, easy-use user interface.

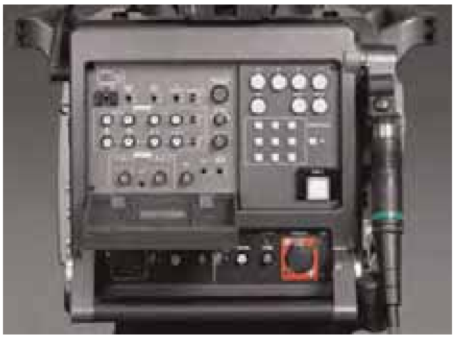


Figure 5 *Control panel of camera head*

6 New functions of CCU

To acquire super high-resolution video and display on a huge panel or screen, preventing images from becoming out of focus is of course essential. To get correct focal adjustment, a video operator with a 4K × 2K resolution monitor controls the focus on the CCU side. For this, we developed new CCU functions (Fig. 6). First, monitoring output can be magnified 'dot-by-dot' using the operating panel. The video engineer can easily select eight areas by eight buttons on the operating panel. Secondly, we developed a serial digital focus control system. The video engineer can control the focus with the focus remote controller attached to the CCU. The serial command between camera and CCU can be multiplexed by one optical cable. Finally, the newly developed camera remote control panel can control many parameters just as a HDTV camera does. It is also possible to use the CCU as a colour corrector after shooting.

7 New functions of video processor

Output from the UHDTV camera is mapped on dual green SDI composed of a 16 HD-SDI (SMPTE-292M). The connection between the camera and the video processor or

Figure 6 *UHDTV CCU*

Figure 7 *UHDTV video processor*

between the video processor and the compressed recorder has an alert system that indicates when the physical connection is incorrect by using a payload ID (Fig. 7). It is possible to monitor HD-SDI error, too. In addition, the video engineer is able to monitor the luminescence of the four G1, G2, B and R signals or the elective colour channel on waveform monitor (WFM) at the same time to manage the two green channels.

4K resolution monitor output, downconverted from 8K × 4K video, uses 3G-SDI (level B). HD monitor output, downconverted from 8K, uses HD-SDI. They can be used for adjustment of 8K output.

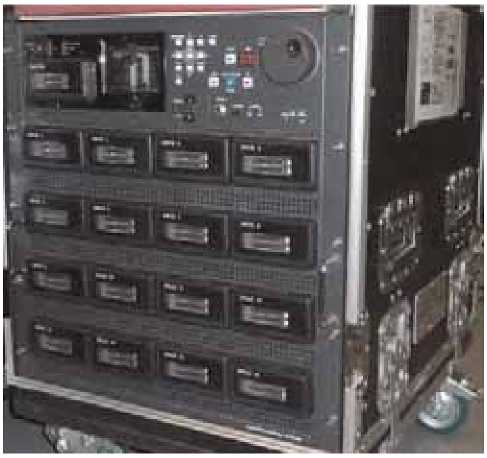


Figure 8 *UHDTV compressed processor*

8 Compressed recorder

We developed for the first time a compression recorder for UHDTV (Fig. 8). Fig. 9 shows a simple configuration. This compressed recorder uses 17 P2 [3] cards to record. 16 out of the 17 slots are for the UHDTV main output. One is for a proxy file. One slot can accept two P2 cards. Total recordable time is about two hours with 34 cards. Output signals are not only the main video, but also 24 AES/EBU channels, downconverted HD-SDI, and stereo analogue audio. The panel is also equipped with a 3.5 inch LCD monitor.

The dual green UHDTV signal that is inputted is separated into 16 streams and compressed with AVC-I codec. The total bit rate for compressed data will be approximately 1.6 Gbits (each P2 card can record a 100 Mbits file). Audio inputs, selectable from HD-SDI embedded or AES/EBU, are recorded in a PCM format (48 kHz 24 bit). RAW data on ANC is stored on P2 cards as uncompressed data simultaneously.

A proxy file created simultaneously is very useful as 8K monitoring is difficult. Video downconverted from UHDTV, and 8ch audio chosen from 24ch and RAW data converted to text memo are compressed internally and recorded on a single P2 card (Table 2).

9 Pre-knee

First, 12 bit sensor output goes through the DSP on the camera head to adjust gain, pedestal, flare, and shading, and then is converted to a 10 bit output signal. The camera head and CCU are connected by one optical fibre multiplexed 3 10 GHz signal. One of 8 pre-knee curves, D 200, D 400, log, super knee and 4 user curves is applied when a 12 bit signal is converted to 10 bits. The video processor restores the signal from 10 to 12 bits, and manages the correction of chromatic aberration owing to the lens, DTL, knee, w. clip and gamma. Finally the dual green SDI signal is formed.

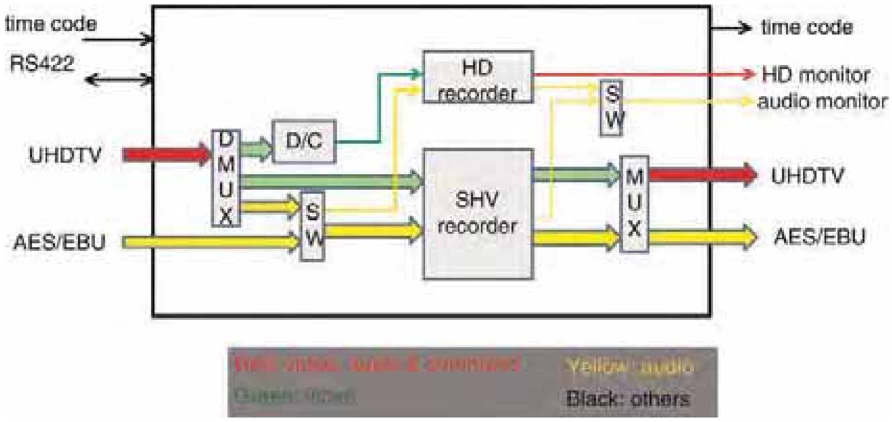


Figure 9 *UHDTV compressed processor-simple configuration*

Table 2 Specifications for compressed recorder

Item	P2 super hi-vision recorder
Input	video: dual green SHV
	audio: AES24ch or SDI embedded (8ch on G1-a, b, c)
Output	video: dual green SHV monitor out (HD D/C) 3.5 inch LCD on front panel (HD)
	audio: AES24ch SDI embedded (8ch on G1-a, b, c) monitor output (analogue), headphone output
Compression	AVC-intra 100 Mbits 16 parallel processing
Medium	flash memory card (P2 card)
	main output 16, Master (HD) 1 total 17
	MAX capacity 64 GB × 34 (17 × 2 sets)
Capacity	138 min (69 min/card)
Proxy	recorded on a P2 card (downconverted)
Size	W440xH487xD500 11 U
	55 kg
Power	330 W
Buck up	LTO tape or HDD
Others	UHDTV RAW data recording

D 200%, meaning 200% of the dynamic range, is suitable for ordinary acquisition with daylight. D 400%, meaning 400% of the dynamic range, has higher sensitivity than D 200% but the S/N is not as good. Log is a film curve that uses the full sensor dynamic range. Super knee is very good for monitoring in the field because it allows gamma and knee curves simultaneously. This makes it possible to recreate highlighted detail (Fig. 10).

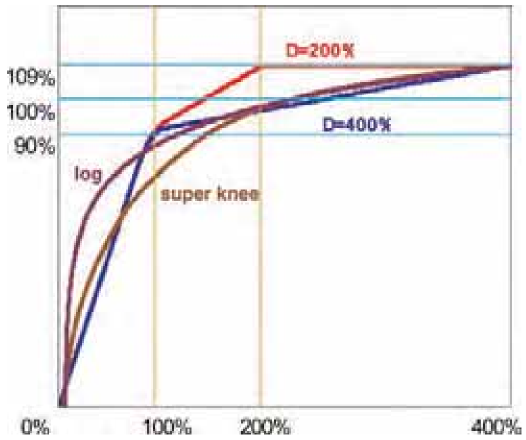


Figure 10 *Pre-knee curve*

10 Raw acquisition

A critical matter for TV shooting is that images are efficiently colour corrected and the post-production does not take such a long time as compared with film shooting. To achieve this efficiency, we devised RAW acquisition capability. With this camera and compressed recorder, adjustment data are saved as ANC packets in the HD-SDI and digitised as data in the compressed recorder. When the RAW mode is selected on the control panel, adjusted DTL, knee, w. clip and gamma in the video processor are not applied on the video output and it is recorded as ANC data. The compressed recorder stores this ANC packet as data in the MFX. When we proceed to post production, it is possible to recreate all kinds of metadata and modify them as well. Moreover, video without DTL, including high frequency elements, is easy to compress and keeps loss lower.

11 Contents production

Fig. 11 shows UHDTV production configuration. One optical fibre cable connecting the camera head and the CCU can transport the main video, supply power, deliver return video, multichannel audio digital interface (MADI),

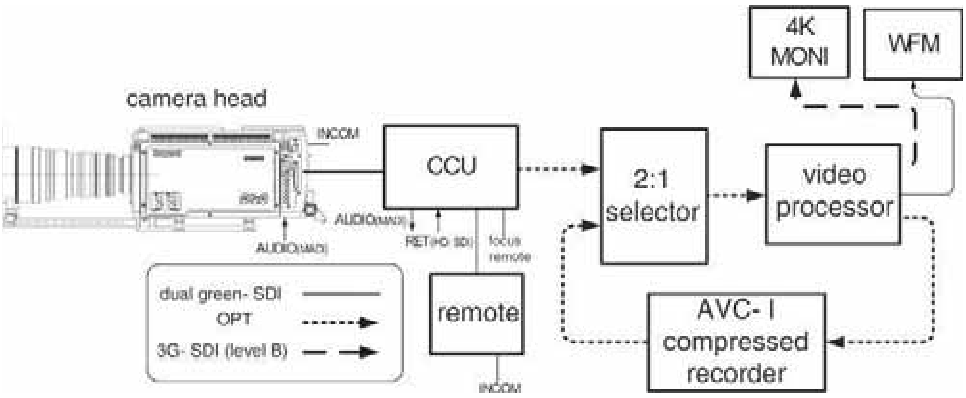


Figure 11 *UHDTV production configuration*

INCOM, and so on. Dual green SDI goes to the 2:1 SW and then to the video processor. The video processor corrects chromatic aberration and adds DTL, gamma, w. clip and knee. Finally they are recorded as compressed data. When RAW mode is selected, no correction is added to the main output, and adjustment data is multiplexed as ANC on HD-SDI. When we use the compressed recorder, the ANC data is stored as data but played out as ANC data. The ANC data can reproduce the actual shot image by using the video processor again. On top of that, ANC data can be modified or changed with the camera remote.

In addition to the camera system, we are developing a real-time RAW implicated UHDTV colour grading system with FPN correction and non-liner editor for compressed UHDTV to edit huge amounts of data efficiently. Moreover, we are also developing a UHDTV live switcher and slow-playback recorder for live sports feeds and a variety of other contents.

12 Conclusions

This latest super hi-vision camera system we have developed is an essential system for supporting many kinds of UHDTV contents production. Moreover, our newly developed RAW acquisition and pre-knee functions allow the camera sensors to fully play out their capabilities and ensure a more efficient production workflow. Furthermore, the development of a switcher and an editorial system will allow the viewer to enjoy a wide variety of high-resolution contents. We will continue to develop this UHDTV technology in cooperation with our R&D Institute to realise UHDTV broadcasting.

13 Acknowledgment

The authors would like to thank the following companies that collaborated with them: Ikegami Tsushinki, Astrodesign, Aptina, Fujinon and Panasonic.

14 References

[1] SMPTE 2036-1-2009: ‘Ultra high definition television – image parameter values for program production’, 2009

[2] YAMASHITA T., HUANG S., FUNATSU R., *ET AL.*: ‘Experimental colour video capturing equipment with three 33-megapixel CMOS image sensors’, *Proc. SPIE*, 2009, **7249**, article id. 72490H

[3] Broadcast and Professional AV Global Web Site, http://panasonic.biz/sav/p2/index.html

Selected content from IBC 2010

ISSN 2041-5923

Step into the light – the EBU loudness recommendation R128

F. Camerer

ORF – Austrian Broadcasting Corporation, Wuerzburggasse 30, A-1136 Vienna, Austria
E-mail: florian.camerer@orf.at

Abstract: The EBU group PLOUD has reached the final stage of its work resulted in recommendations and documents that will have a profound effect on any audio production in broadcasting. The gradual switch from peak to loudness normalisation combined with a new maximum true peak level and the usage of the descriptor 'loudness range' allow for the first time to fully characterise the audio part of a program. More importantly it has the potential to solve the most frequent complaint of the listeners, that of severe level jumps within and between programmes. At the core of PLOUD's output stands the loudness recommendation R128. The origin of the issue as well as the details of the recommendation will be presented.

1 Introduction

In Europe, there still exists an audio metering and levelling paradigm based on the so-called the 'quasi peak programme meter' (QPPM). It is 'quasi' because of its finite reaction time, which is 10 ms (although 5 ms also is also found these days).

In practice, this means that signal peaks shorter than this reaction time won't be displayed correctly, if at all (for example, transients – think of jangling keys). The agreed permitted maximum level (PML) now is -9 dBFS, measured with bespoke QPPM. That level was related to the modulation of a TV channel (FM modulation) in order to provide headroom for those transients that one would not see on the meter, but which should nevertheless be there so as to contribute to the 'openness' of the audio signal. This relationship was provided by aligning -9 dBFS PML to 30 kHz deviation on the FM carrier (not too long ago, the most widely used analogue transmission system). The maximum deviation allowed is 50 kHz (for TV channels), so that gave a 'headroom' of 20 kHz or 4.4 dB (see Fig. 1).

With the more widespread switch to digital production and the use of more sophisticated audio processors someone 'discovered' that if you aligned your PML of -9 dBFS to the maximum allowed deviation (50 kHz) you would get a 4.4 dB louder signal on air – but, you had to radically cut off those short transients above -9 dBFS, as they would cause overmodulation and thus lead to all sorts of nasty artefacts like excessive sibilance in speech. Once this was discovered – and put into practice – the loudness war began. Commercial broadcasters were the first, as 'louder is always better', and the public ones reluctantly had to follow. Better and better multiband processors were used to push the loudness level close to the allowed peak level, leading to overcompressed sound, listener fatigue, loss of transparency and generally the situation so many complain about today.

1.1 First step towards a solution – the ITU loudness metering recommendation

Many in our industry thought that this had turned into a hopeless situation, only surpassed by the music business, where loudness levels of, especially, pop-music CDs have reached absolutely ridiculous heights. Miraculously, light at the end of the tunnel appeared in 2006 with the international standardisation of a way to measure loudness. The work and tests have been conducted by the ITU, the International Telecommunications Union, and resulted in ITU-R BS.1770: 'Algorithms to measure audio programme loudness and true-peak audio level' [1]. This breakthrough generated a new wave of loudness awareness and work; it was also one of the inspiring factors for the birth of PLOUD.

Briefly, ITU-R BS.1770 provides a simple and practical solution for the task of finding an objective measurement of

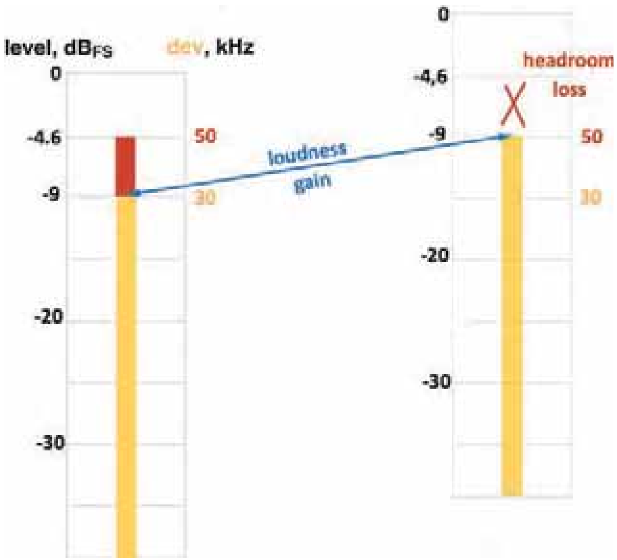


Figure 1 *Relationship between permitted maximum level (with QPPM) and FM deviation and the abuse of it, gaining loudness, but losing headroom*

what is essentially a subjective impression (loudness). Such a measure can naturally never be perfect, as the perception of loudness depends on many parameters, such as preference, age, mood, frequency, replay level etc. Nevertheless, BS.1770 proved to be a robust standard with the added value of simple implementation.

ITU-R BS.1770 basically consists of a weighting curve (similar to the 'B-weighting' curve) called 'K-weighting', a second order high-pass filter with an additional high-frequency-shelving filter (see Fig. 2).

The mean-square energy of all K-weighted channels (except the low frequency effects (LFE) channel which is not included in the measurement) is now calculated with a certain gain factor (0 dB for the 3 front channels, +1.5 dB for the surround channels of a multichannel audio signal), and the result is displayed as LKFS (loudness, K-weighting, referenced to digital full scale) (see Figs. 3 and 4). For relative measurements, loudness units (LU) are used, where 1 LU is equivalent to 1 dB.

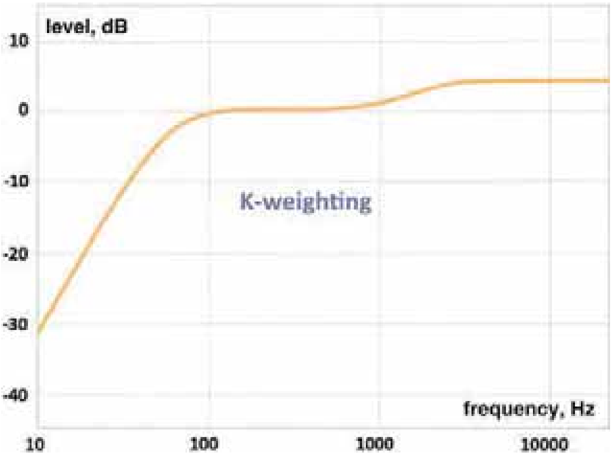


Figure 2 *K-weighting curve for loudness measurement*

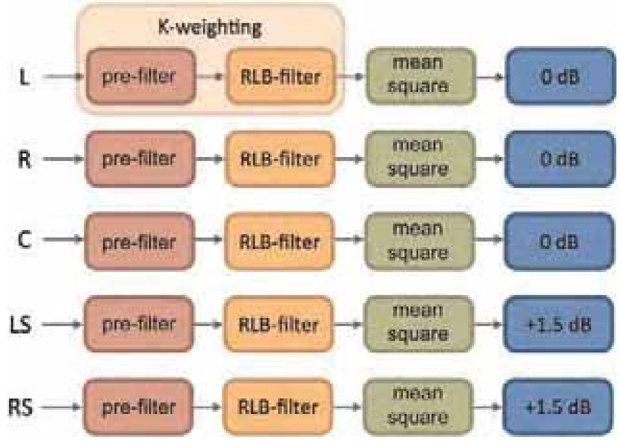


Figure 3 *Channel processing in ITU-R BS.1770*

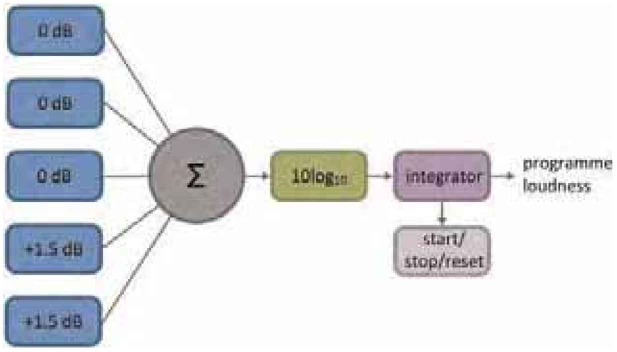


Figure 4 *Channel summation in ITU-R BS.1770*

As can be seen from the title of ITU-R BS.1770, the recommendation also specifies a way to accurately measure peaks – with an oversampling true peak meter (dBTP; see also later). This catches inter-sample peaks that potentially reach beyond 0 dBFS, that could cause distortion in D-to-A-converters and low bit rate codecs.

2 Ploud gets loud

With ITU-R BS.1770 in place, PLOUD set out to radically change the audio levelling and metering paradigm: abandoning QPPM and peak normalisation, and embracing loudness metering and normalisation.

It cannot be emphasised enough that the adoption of the loudness levelling paradigm will be a fundamental change for the audio industry, first of all for the broadcast industry, but in the mid-term future for the music recording and mastering business as well. It must also be stressed that the scope of PLOUD's work is the whole transmission chain – production, distribution and transmission – to guarantee consistent levels throughout facilities, at distribution companies, rebroadcasters and ultimately for the listener at home. The days of having to regularly use the remote control to adjust widely varying audio levels are coming to an end! That at least is the big goal PLOUD is aiming at.

At the heart of the forthcoming EBU recommendation R128 ‘Loudness normalisation and permitted maximum level of audio signals’ lies the substitution of QPPM levelling with 3 new audio parameters or descriptors:

- Programme loudness
- Loudness range (LRA) and
- True peak level (TPL)

Let us now examine those 3 parameters in detail.

2.1 Programme loudness, target level

This is arguably the most important parameter and describes the long-term integrated loudness over the duration of a programme or complete item. The parameter consists of one number (in LKFS, with one number after the decimal point) that indicates ‘how loud the programme is on average’. This is measured with a meter compliant to ITU-R BS.1770 with the addition of a gating function. The gate serves to pause the loudness measurement when the signal drops below a certain threshold. Programmes with a considerable amount of low-level audio will therefore have their loudness determined by the ‘foreground sounds’, which results in better alignment with other content which has less low-level audio.

PLOUD conducted listening tests in Q4/2009 and January 2010 to determine the best gating threshold. It was found that two candidate gating methods out of the four tested gave good results, both being statistically significantly better than the other two.

Those two methods were a gate of 6 LU relative to ungated LKFS (‘ 6rel’) and 10 LU relative to ungated LKFS (‘ 10rel’). For all candidates, a block length of 400 ms was used. Pragmatically, a value of 8rel was chosen for further informal tests against the other gating function already used by broadcasters: ‘Dialogue Intelligence’, a proprietary algorithm of the Dolby company. The difference in loudness between the Dolby algorithm and the 8rel gate was less than 1 LU for programmes with little dynamic content (see loudness range descriptor later); for content with a greater loudness range, the difference was 1-2 LU on average.

Checks were also performed on the original ITU database to see if there were any significant effects using the 8rel gate. As expected, this was not the case. A gate of 8 LU relative to ungated LKFS therefore seems to be a robust algorithm that is likely to achieve the goal of satisfactorily aligning wide-loudness-range programmes with programmes which exhibit a narrower loudness range. As it is a neutral method, not dependent on any signal type (such as music, voice or sound effects), this gate cannot be tricked. The values obtained with 8rel are close enough to dialogue gating, so that broadcasters who already use that method can make an easy transition.

What is now the ‘magic number’, one will ask, the ‘new centre of the audio universe’?

A lot of thought went into the answer to this fundamental question. The distribution subgroup of PLOUD addressed it in detail. A main consideration was the necessary loudness range to fit in all programmes (provided that they don’t exceed the tolerable loudness range for domestic listening).

The magic number is

$$-23 \text{ LKFS } (-8\text{rel gate})$$

This is the so-called ‘target level’ to which every audio signal will be normalised (a deviation of ± 1 LU is acceptable for programmes that are transmitted live or with a similar time constraint; in the transitional phase, slightly more deviation might occur). To put it in perspective: 9 dBFS (measured with a QPPM) is no longer the value we normalise to, but 23 LKFS (or LUFS, see later). Quasi-peaks are losing their foothold.

In systems where loudness metadata can be set (e.g. the parameter ‘dialnorm’ in the Dolby AC-3 system) this metadata parameter shall also be set to 23, when the content has been normalised to that value. If it has not been normalised, then loudness metadata shall in all cases correctly indicate the loudness level. This enables us to keep current mixing practices for the transition to full loudness normalisation, including in production – but the aim certainly is to encourage 23 as the common standard for all programmes.

In the transitional phase, it might also be reasonable to install loudness processors that adjust the level of e.g. legacy material that is not yet compliant to 23. Archive material will benefit especially from this approach, as it will take some time after the decision is taken, to loudness normalise all stored content.

2.2 Loudness range (LRA)

The loudness range descriptor quantifies (in LU) the variation of the loudness measurement of a programme. It is based on the statistical distribution of loudness within a programme, thereby excluding the extremes. Therefore, for example, a single gunshot is not able to bias the outcome of the LRA computation. This is very similar to what statisticians call the ‘percentile range’. The algorithm for LRA has kindly been provided by the company TC

Electronic, based on their years of work and experience with this subject.

The upcoming EBU recommendation R128 will not specify a maximum permitted LRA, as this is dependent on factors such as the tolerance window of your average listener's habits, the distribution of genres of your station etc. R128 does, however, strongly encourage you to use LRA to determine if dynamic treatment of an audio signal is needed and to match the signal with the requirements of a particular transmission channel or platform. Loudness range is a useful indicator to decide when and how much of such processing might be applied. Fig. 5 shows the loudness distribution and LRA of the movie 'The Matrix'; 25 LU is probably challenging for most living rooms.

Consequently, first experiences at broadcasting stations suggest a maximum LRA value of approximately 20 LU for highly dynamic material, such as action movies or classical music. The majority of programming will never need to fully use such a high LRA value or, indeed, be able to reach it! If programmes (think of a Bruckner symphony or many action movies) exceed this value, one might then carefully reduce their loudness range or decide to even leave it untouched, if your target audience, genre portfolio and the transmission characteristics of your channel make it seem reasonable.

2.3 True peak level (TPL), maximum TPL

The true peak level indicates the maximum (positive or negative) value of the signal waveform in the continuous time domain; this value may be higher than the largest sample value in the time-sampled domain. With an oversampling true peak meter compliant to ITU-R BS.1770, those true peaks are now able to be detected (the accuracy depends on the oversampling frequency). It is only necessary to leave a headroom of 1 dB below 0 dBFS to still accommodate the potential under-read of about 0.5 dB (for a 4x oversampling true-peak meter; basic sampe rate: 48 kHz).

The maximum true peak level recommended in R128 will consequently be

–1 dBTP

This is applicable to the production environment. For legacy analogue re-broadcasters, for example, a lower true peak level is needed as a result of the loudness normalisation paradigm and the relationship to the permitted FM deviation. Also low bit rate codecs might need more headroom. The PLOUD Distribution Guidelines will address these and other issues in detail.

3 Metering

A further subgroup within PLOUD, consisting of some of the familiar names of audio metering equipment, has come up with suggestions for the basic properties of an R128-compliant loudness meter. Those suggestions are based on continued discussions concerning the various parameters that would be needed in a practical metering device. The results will provide the general framework for a so-called 'EBU mode' of a loudness meter. Some of the specifications of the EBU mode will be as follows

3.1 Time constants

Three different integration times for three 'flavours' of loudness will be used

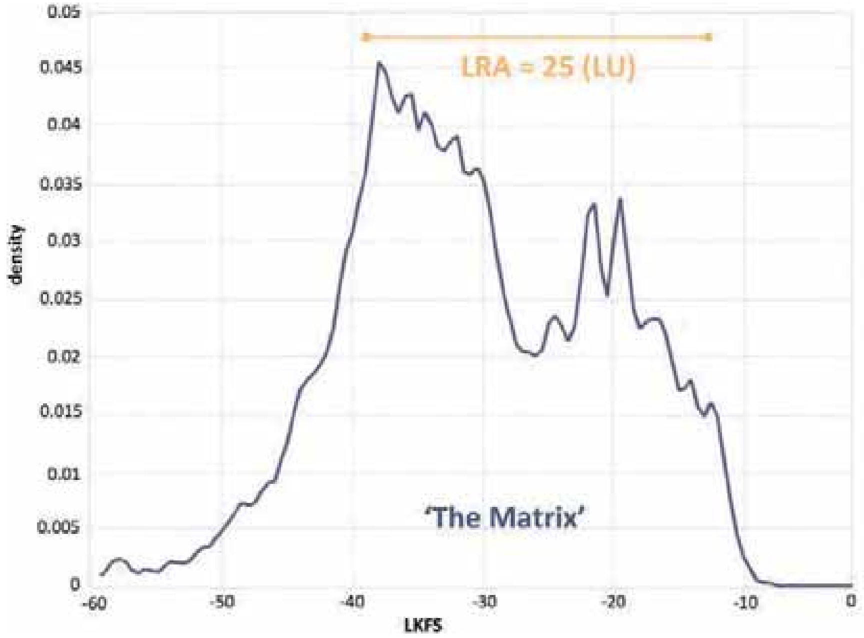


Figure 5 *Loudness range (LRA) as a result of the statistical distribution of loudness*

- 400 ms: Momentary loudness (ungated)

- 3 s: Sliding window loudness (ungated) and

- start/stop: Integrated loudness (with 8rel gate)

In all cases the measurement will be performed as specified in ITU-R BS.1770.

3.2 Units

Owing to an inconsistency between ITU-R BS.1770 and ITU-R BS.1771, a different naming convention is suggested, complying to ISO 80000-8:

- The symbol for 'loudness level' should be 'L_K'.

- The unit symbol 'LUFS' indicates the value of L_K with reference to digital full scale.

- The unit symbol 'LU' indicates the L_K without an absolute reference and thus also loudness level differences.

3.3 Scale and range

The scale is based on the target level of 23 LUFS. The basic scale covers a range of 27 LU, from 18 to +9 LU relative to the target level (or in absolute terms: 41 to 14 LUFS). The extended scale needed for programmes with a wide loudness range doubles the basic range, covering 54 LU, from 36 to +18 LU (59 to 5 LUFS). '0 LU' equals the target level of 23 LUFS (see Fig. 6).

The suggested scales also try to mimic the appearance of some legacy peak meters with respect to a larger part (two thirds) lying below the zero-point, thereby providing some familiarity and enabling an easier switch to the new metering method.

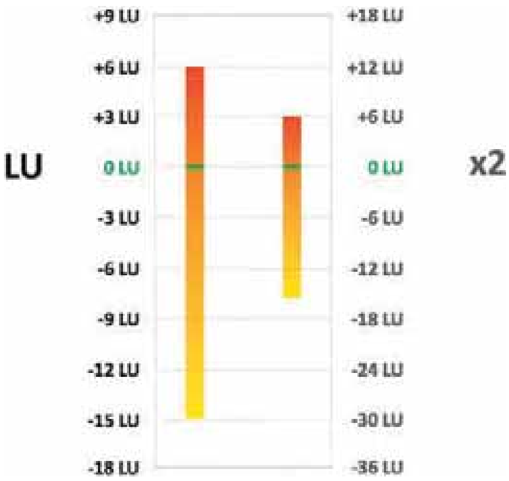


Figure 6 *Loudness metering scales in LU; basic and extended version*

3.4 Safety gate

A safety gate with an absolute threshold of 70 LUFS will be used to enable a loudness measurement to start exactly when the audio signal starts.

4 Practical guidelines, distribution guidelines

As the actual recommendation R128 will be short and concise, it is advantageous to explain its consequences in a separate document. The Practical Guidelines serve that purpose. The concepts of R128 will be explained in depth, giving guidance to mixing engineers, planning and implementation departments, coping with metadata and delivery specifications, alignment strategies and other studio issues.

A very active subgroup is covering all aspects of loudness normalisation for the distribution of audio signals, addressing the critical links between production and the final recipient, the consumer. The scope of the subgroup's work includes active compensation of sources that differ from the R128 loudness target level, without affecting the creative properties of the content. This is another significant step to eliminate the motive for loudness competition between channels. The level relationships between the different physical outputs of set-top-boxes (SPDIF, HDMI, analogue RCA, SCART etc.), the required maximum loudness-to-peak ranges, metadata issues and a blueprint for loudness-compliant design of the audio interfaces of consumer equipment are other topics on the agenda. The IT Distribution Guidelines fulfil a crucial role in harmonising the whole signal chain right to the recipient with respect to loudness normalisation.

5 Conclusions

The future practical implementation of the work of PLOUD will fundamentally affect the way we treat audio levels. Switching to loudness metering and normalisation is a breakthrough in providing a more consistent, and thus more pleasant, listening experience for the consumers.

PLOUD will issue five documents:

- the actual loudness recommendation R128

- a Metering Technical Document

- a Loudness Range Technical Document

- Practical Guidelines for implementation and production and

- Distribution Guidelines

Publishing these documents can nevertheless only be the first step.

The active promotion, widespread adoption, implementation and realisation of the concepts within these documents will be the challenge ahead on the road to an international loudness levelling solution.

6 Reference

[1] ITU-R BS.1770-1: Algorithms to measure audio programme loudness and true-peak audio level, 2006–2007

Selected content from IBC 2010

ISSN 2041-5923

Wireless and fibre-optic live contribution link for uncompressed super hi-vision signals

T. Nakatogawa S. Okabe M. Nakamura K. Oyamada F. Suginoshita T. Ikeda K. Shogen

Science & Technology Research Laboratories, NHK (Japan Broadcasting Corporation), 1-10-11 Kinuta Setagaya-ku, Tokyo 157-8510, Japan
E-mail: nakatogawa.t-iq@nhk.or.jp

Abstract: Super hi-vision (SHV) is a type of ultra-high definition television (HDTV) system. This paper describes live contribution technologies for a dual-green format SHV, whose uncompressed signal consists of 16 high definition serial digital interface (HD-SDI) signals having a total bit rate of 24 Gbit/s. For short-haul wireless links, three pairs of 10.3 Gbit/s transmitters and receivers operating in the 120 GHz band are used to transmit an uncompressed SHV signal. Each 10.3 Gbit/s signal contains up to six HD-SDI signals and concatenated Reed-Solomon forward error correction codes. For long-haul fibre-optic links, a converter was developed for changing an uncompressed SHV signal into an optical transport unit 3 (OTU3) signal, commonly used in 40 Gbit/s core/metro networks. An outdoor trial of a wireless link and an indoor trial of a fibre-optic link achieved quasi error-free transmission of a SHV picture without interruption.

1 Introduction

Super hi-vision (SHV) is a type of ultra-high definition television (HDTV) system with four times as many lines and pixels per line as HDTV. SHV is recommended in Rec. ITU-R BT.1769 [1] for extended large screen digital imagery (LSDI) systems and also in SMPTE 2036-1-2007 [2] for ultra-HDTV. A dual-green SHV format with a pixel offset and combining four 8 million pixel sensors/panels (R, G1, G2 and B) has been developed [3]. The uncompressed signal consists of sixteen 1.5 Gbit/s high definition serial digital interface (HD-SDI) signals having a total bit rate of 24 Gbit/s.

Live contribution links for SHV are being developed. While international contribution links for SHV signals were demonstrated at IBC 2008, the received pictures were delayed by the video codec that was used [4]. Video codec technologies supply broadcasters with cost-effective and feasible contribution links. However, the contribution links used especially for live broadcasts must be capable of stably carrying video and audio with minimal latency. By making it possible to transmit uncompressed signals from various remote locations, the contribution link system permits live switching between studio and remotely-located cameras.

This paper describes short-haul wireless and long-haul fibre-optic technologies for live transmission of an uncompressed dual-green format SHV signal, as shown in Fig. 1. In the following sections, we propose two links for the SHV transmission described in Fig. 2.

2 Wireless contribution link

A wireless transmission system for live broadcasts that can transmit an uncompressed HD-SDI signal output from an HD camera using 60 GHz band radio waves has already been developed and used in some events such as the Torino 2006 Winter Olympic Games [5].

The authors are developing a wireless contribution link, called an field pickup unit (FPU), intended for use in the production of live sports and other programs shot with multiple HD cameras. The FPU in development operates in the 120 GHz band and enables transmission of up to six multiplexed HD-SDI signals as a 10.3 Gbit/s signal [6].

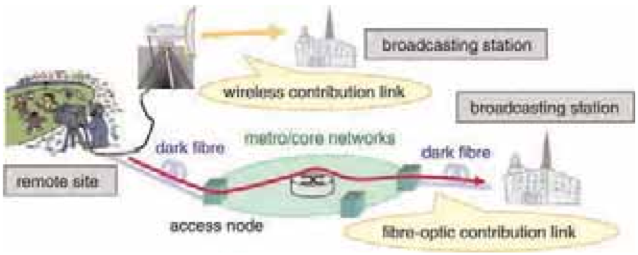


Figure 1 *Contribution links to broadcasting stations*

Here, we report on wireless transmission experiments of an uncompressed SHV signal using three sets of 120 GHz band wireless transmission equipment.

2.1 Overview of the 120 GHz band wireless transmission system

The specifications for the 120 GHz band wireless transmission system are presented in Table 1. We developed a stable, low latency error correction system to handle six multiplexed HD-SDI signals. Reed-Solomon (RS) code is used for error correction because it makes possible high-speed encoding and decoding by means of parallel processing. To improve error correction performance, we use a method of concatenated RS codes that yields an encoding gain of about 6 dB at a bit error ratio (BER) of 1×10^{10}. Our experiments involved transmission of a 10.3 Gbit/s signal composed of six multiplexed HD-SDI signals over a 3 km distance, and it showed that quasi error-free transmission was possible in clear weather [6].

The uncompressed SHV signal is configured as six, six and four HD-SDI signals for transmission by the respective 120 GHz band wireless units. Only one channel in the 120 GHz band is available for our current equipment; thus, all of the signals are transmitted at the same frequency.

The following sections describe the effects of co-frequency interference between three 120 GHz signals.

2.2 Short-haul transmission trial

2.2.1 Experimental overview:

Short-haul links will be used for transmissions of 100 m or less in places where it is difficult to lay cable, such as from buildings and over rivers and roads. For buildings, we also assume that the transmissions will be through window glass. We therefore conducted transmission experiments over distances of 40 and 60 m through glass. Fig. 3 shows the setup of the experiment, and Fig. 4 shows the three receiving antennas. The separations between the transmitting antennas and between the receiving antennas (denoted by 'd' in Fig. 3) were 1, 3 or 5 m. The three wireless transmission paths were parallel. The criterion for quasi error-free transmission was 1×10^{10} or less.

Table 1 Specifications of the 120 GHz band wireless transmission system

Centre frequency		125 GHz
Modulation scheme		amplitude shift keying
Transmission output power		10 mW
Frequency bandwidth		17 GHz
Transmission bit rate		10.3 Gbit/s
Error correction	outer code	RS (986, 966)
	inner code	RS (252, 236)

2.2.2 Experimental results:

With error correction, the BER results were 1×10^{10} or less for all conditions except d of 1 m. The results showed that interference between parallel transmissions over a short distance affects receiving quality and that stable transmissions can be achieved by using error correction. There were no problems with the received SHV video even though the transmissions passed through window glass. Fig. 5 shows the equipment used for monitoring the received SHV video.

2.3 Trial transmission over more than 1 km

2.3.1 Experimental overview:

We conducted experiments with a transmission distance of 1.25 km. Fig. 6 shows the setup of the experiments. Transmitters and receivers were set on roofs of buildings and the experiments were conducted under the following conditions:

Condition 1: The three transmitters and receivers used vertical polarisation.

Condition 2: The centre transmitter and receiver (Tx2 and Rx2) used horizontal polarisation. The others used vertical polarisation.

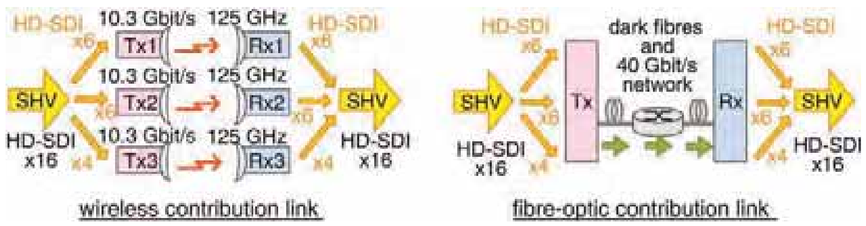


Figure 2 *Proposed transmission systems for uncompressed SHV signal*

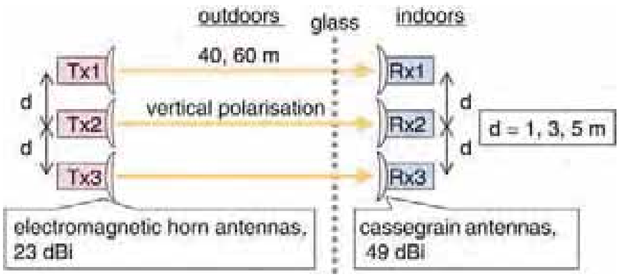


Figure 3 *Setup of short-haul experiment*

Figure 4 *Receiving antennas*

Figure 5 *Received SHV video*

2.3.2 Experimental results: Under Condition 1, co-frequency interference caused a large degradation in the received signal even when the receivers were 8 m apart; the receiver's frame-synchronisation was lost, and there was no reception at all.

Under Condition 2, error correction resulted in a BER of 1×10^{10} or less for all three receivers. Furthermore, even when Rx2 with horizontal polarisation and Rx3 with vertical polarization were placed immediately beside each other, as shown in Fig. 7, error correction resulted in a BER of 1×10^{10} or less. There were no problems with the received SHV video under Condition 2.

3 Fibre-optic contribution link

There are two means of constructing stable fibre-optic links for uncompressed SHV signals. One is to use dark fibre, which is unused optical fibre available for use; the other is to use lease lines of core/metro networks. If a link utilises only dark fibres, it is possible to transmit an SHV signal simply by only carrying out electrical to optical (E/O) conversion of each of the 16 HD-SDI signals and wavelength division multiplexing (WDM) of the converted 16 HD-SDI optical signals. However, since a long-haul link that utilizes only dark fibres is expensive, it would be more feasible to construct a hybrid link that utilises both dark fibres and lease lines. In that case, the SHV signal has to be converted into a signal with a lease line format for its transmission.

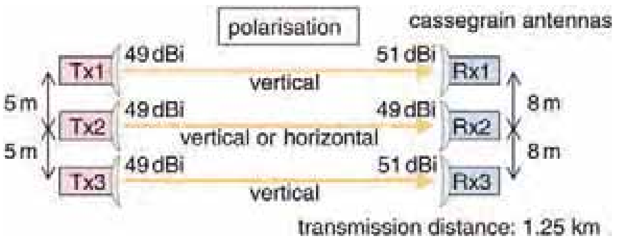


Figure 6 *Setup of 1.25 km transmission experiment*

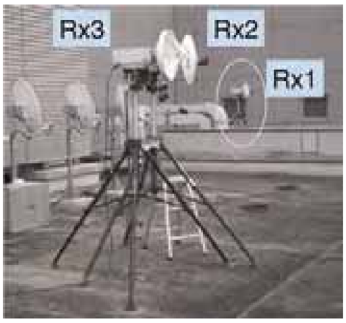


Figure 7 *Receiving antennas*

More and more fibre-optic transmission systems of high-speed core/metro communication networks are becoming optical transport network (OTN), as standardised in ITU-T Rec. G.709. OTN technology is commonly called digital wrapper technology [7]. The line rate of optical transport unit 3 (OTU3) defined in OTN is 43 Gbit/s. It is enough to transmit 24 Gbit/s uncompressed SHV signal together with RS (255, 239) parity bits.

The following sections describe a method of conversion from an uncompressed SHV signal to an OTU3 signal for fibre-optic contribution links and an indoor experiment.

3.1 Conversion of 24 Gbit/s SHV signal into OTU3 signal

Fig. 8 shows the procedure for converting an SHV signal into an OTU3 signal, the payload of which can accommodate up to four 9.95 Gbit/s signals. First, up to six of the 16 HD-SDI signals making up the SHV signal are converted into a 9.95 Gbit/s signal having an synchronous transport module 64 (STM64) frame structure [8]. Next, each 9.95 Gbit/s signal is converted into an optical channel data unit 2 (ODU2) signal [7]. The three ODU2 signals and an ODU2 signal consisting of null data are then put into the

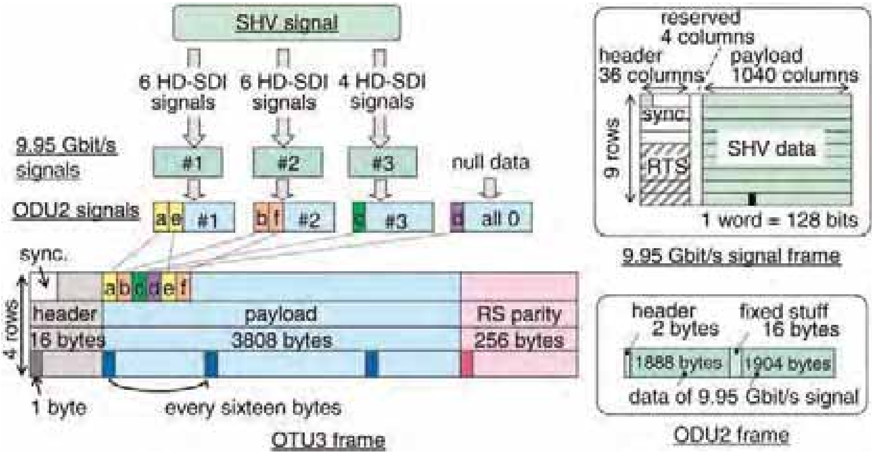


Figure 8 *Procedure of converting SHV signal into OTU3 signal*

payload of an OTU3 signal in accordance with ITU-T Rec. G.709. RS (255, 239) codes are calculated for every sixteen bytes of the header and payload of the OTU3 signal. The OTU3 frame in Fig. 8 is generated in a 3.35 microsecond cycle and transmitted at a bit rate of 40 Gbit/s (at a line rate of 43 Gbit/s including a redundant RS code bit rate).

In our equipment, an field programmable gate array (FPGA), a general-purpose LSI with a reconfigurable signal processing circuit) converted up to six HD-SDI signals into a 9.95 Gbit/s signal, and a commercially supplied LSI for OTNs converted 9.95 Gbit/s signals into an OTU3 signal.

3.2 Transmission of SHV clock signal

An SHV clock signal, equal to each HD-SDI clock signal, must be recovered from an OTU3 signal received at the receiver site. Since the SHV clock signal and OTU3 core/metro network clock signal are asynchronous, the link requires an accurate clock recovery method. Our system uses the synchronous residual time stamp (SRTS) method, standardised in ITU-T Rec. I.363.1 [9], to convert the SHV 74.25 MHz clock signal into a 77.76 MHz clock signal, which is synchronous with an OTU3 signal.

Fig. 9 illustrates the proposed clock recovery method utilising SRTS. At the transmitter, the number of cycles of a 77.76 MHz clock signal (f_n) within N cycles of the SHV 74.25 MHz clock signal (f_s) is counted and transmitted to a receiver as a variable residual time stamp (RTS). At the receiver, the counter circuit is driven at the cycle of the 77.76 MHz clock signal recovered from the received OTU3 signal. Short pulses are output only if the counter value is equal to the received RTS value. Since each Interval between pulses is almost equal to N cycles of the SHV 74.25 MHz clock signal, the SHV 74.25 MHz clock signal is generated by multiplying N to a series of pulse signals. In our equipment, the RTS values were put into the header of each 9.95 Gbit/s signal, and the N value was set to 21.

3.3 Indoor experiment

3.3.1 Jitter characteristics:

Jitter characteristics were measured to evaluate the SRTS method for SHV clock signal transmission. Since the jitter specifications of an

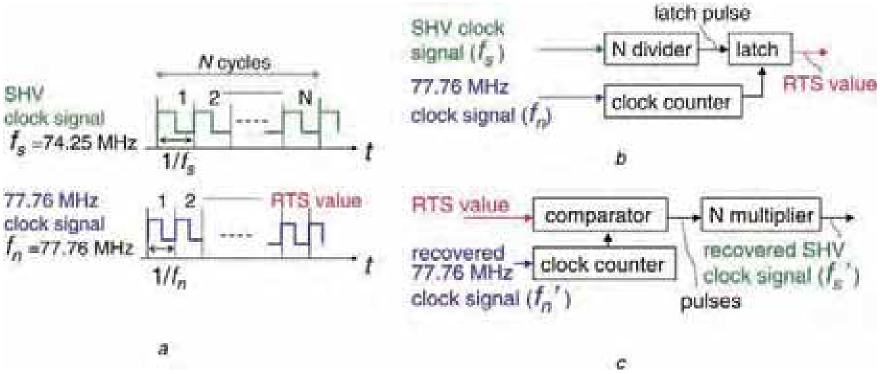


Figure 9 *Clock recovery method using SRTS*

a Clock signal wave forms
b RTS generation at transmitter
c SHV clock recovery at receiver

SHV signal are not standardised at this moment, the jitter characteristics of HD-SDI signals were measured instead. The required characteristics of the HD-SDI alignment jitter (the lower band edge is 100 kHz) and timing jitter (the lower band edge is 10 Hz) as standardised in SMPTE 292-2008 [10] are 0.2 unit interval (UI) or less and 1 UI or less, respectively.

In the indoor trial, HD-SDI colour bar signals were converted into an asynchronous 40 Gbit/s OTU3 signal and transmitted over a 2 km single mode optical fibre. The alignment jitter characteristic and timing jitter characteristic of the received signal were 0.12 and 0.20 UI, respectively, which met the jitter requirements.

3.3.2 BER characteristics: The BER characteristics of the received SHV signal were measured in order to evaluate the method of conversion from an SHV signal to an OTU3 signal. Fig. 10 shows BERs for equipment utilizing optical amplitude shift keying modulation, which is equivalent to on-off keying modulation, whose wavelength and output optical power were 1.55 micrometres and 0 dBm, respectively. The equipment was capable of quasi error-free transmissions over 5 km, which means the BER was 1×10^{10} or less, without error correction and quasi error-free transmission over 6 km with error correction.

Furthermore, the equipment was capable of 50 km quasi error-free transmission without error correction when utilising return-to-zero differential phase shift keying (RZ-DQPSK) modulation, commonly used in commercial long-haul OTU3 core networks, two optical amplifiers, and a dispersion compensation fibre.

3.3.3 Latency characteristics and an SHV video transmission: A latency time of the whole signal processing of the transmitting and receiving equipment was 245 microseconds, which is small enough compared with the duration of a video frame. We demonstrated quasi error-free transmission of an SHV video without interruption over the course of the experiment, which was about a day.

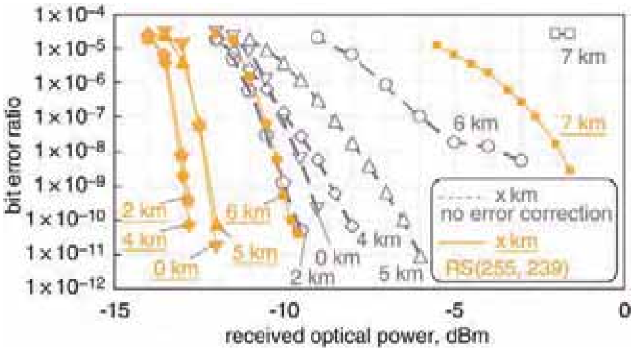


Figure 10 *BER characteristics of the received SHV signal*

These results show that the proposed system is capable of transmitting uncompressed SHV signals over core/metro networks. It would be feasible to transmit SHV over a hybrid link that utilises both laid dark fibres and commercial lease lines.

4 Conclusions

This paper describes short-haul wireless and long-haul fibre-optic technologies for live transmission of a dual-green format uncompressed SHV signal consisting of 16 HD-SDI signals having a total bit rate of 24 Gbit/s.

For short-haul wireless links, three 10.3 Gbit/s transmitter and receiver pairs operating in the 120 GHz band were developed to transmit an uncompressed SHV signal. Although the same 120 GHz band of radio frequencies was used by all of the transmitters, stable transmission over more than 1 km was achieved by using error correction to counter the effects of co-frequency interference.

For long-haul fibre-optic links, we developed a converter for changing an uncompressed SHV signal into an OTU3 signal, and it achieved quasi error-free transmission of a SHV picture without interruption in an indoor experiment. Although we still have to do a test in actual 40 Gbit/s OTU3 networks, the results of the indoor experiments show that the proposed system can transmit SHV signals over core/metro networks.

These new contribution links make possible live SHV production at various locations.

5 Acknowledgments

Part of this research about the fibre-optic link was supported by the New Energy and Industrial Technology Development Organization, Japan, under a project entitled ‘The Development of Next-generation High-efficiency Network Device Technology’.

Part of this research about the wireless link was supported by the Ministry of Internal Affairs and Communications, Japan, under a project entitled ‘The Research and Development Project for the Expansion of Radio Spectrum Resources’.

The authors would like to thank NTT Microsystem Integration Laboratories for developing the 120 GHz wireless equipment.

6 References

[1] Recommendation ITU-R BT.1769: ‘Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange’, 2006

[2] SMPTE 2036-1-2007: ‘Ultra High Definition Television - Image Parameter Values for Program Production’, 2007

[3] SHIMAMOTO H., YAMASHITA T., KOGA N., *ET AL.*: ‘An 8k × 4k ultrahigh-definition color video camera with 8M-pixel CMOS imager’, *SMPTE Motion Imaging Journal*, 2005, **114**, (7–8), pp. 260–268

[4] ZUBRZYCKI J., DAVIES T., SMITH P. C., *ET AL.*: ‘Super hi-vision—the London-Amsterdam live contribution link’. EBU Technical Review, January 2009, pp. 17–34

[5] IZUMOTO T., UMEDA S., OKABE S., *ET AL.*: ‘Pseudo no-delay HDTV transmission system using a 60 GHz band for the Torino Olympic Games’. Proc. of 2006 Int. Broadcasting Convention, 2006, pp. 89–94

[6] OKABE S., IKEDA T., SUGINOSHITA F., *ET AL.*: ‘10-Gbps forward error correction system for 120-GHz-Band wireless transmission’. Proc. of the Fifth Ann. IEEE Radio & Wireless Symp., WE2C-3, 2010, pp. 472–475

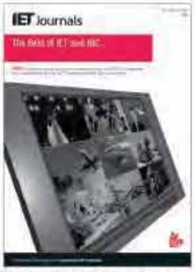
[7] ITU-T Recommendation G.709/Y.1331: ‘Interfaces for the optical transport network (OTN)’, 2009

[8] ITU-T Recommendation G.707/Y.1322: ‘Network node interface for the synchronous digital hierarchy (SDH)’, 2007

[9] ITU-T Recommendation I.363.1: ‘B-ISDN ATM adaptation layer specification: Type 1 AAL’, 1996

[10] SMPTE 292-2008. 2008. 1.5 Gb/s Signal/Data Serial Interface

Selected content from IBC 2010



ISSN 2041-5923

Interactive visualisation of live events using 3D models and video textures

B.A. Weir   G.A. Thomas

BBC R&D, Oxford Road, Manchester M60 7HB, UK
E-mail: bruce.weir@bbc.co.uk

Abstract: For some years, the BBC has been experimenting with technology and algorithms for blending rendered 3D graphics with live video to create television programmes that would previously have been either impossible or too expensive to create using existing technology. In this paper, the authors describe an evolution of these techniques that allows the creation and delivery of live, interactive, 3D models of events that can be viewed, in real-time, by a domestic user via a standard web browser. These events could include sporting fixtures (such as the Olympic Games), dramas or reconstructions for news programmes. The authors describe the system from image capture, through image processing and object detection, recognition and segmentation to data amalgamation, compression, delivery and display. A number of technical advances are described including rapid calibration of camera position and pose, and real-time markerless (and sensorless) tracking of camera motion.

1 Introduction

Although there have been many successful television programmes, such as as the BBC's Bamzooki programme [1] or the ITN news [2], that have mixed rendered 3D graphics and video in a live, or near-live, environment using virtual studio or augmented reality technologies, the output of the graphics systems have always been used as inserts in a traditional, linear television broadcast. The increase of domestic internet speeds, the popularity of IP-based video delivery systems, such as BBC iPlayer, and the general improvement in processing power in home PCs have come together to make it possible to supply the output of 3D compositing systems directly to the end user, allowing the freedom to explore the content interactively.

This functionality is of interest if it allows BBC content makers to tell a story or describe an event more engagingly and effectively to our audiences, and may prove particularly useful where events are spread over a geographically wide area. It is difficult to convey the scope of such events purely from camera imagery, as traditional television simply provides a series of windows onto the scene rather than a broad overview. Coverage of events, such as the London Marathon, are often augmented by computer-generated fly-throughs, showing the extent of the course, but up to now the addition of live video elements into such fly-throughs has not taken place; this would seem to offer exciting and engaging possibilities in developing the narrative of the event. It would make a particularly compelling service if the user could control or select the viewpoint on the scene and an Olympic service, with its many simultaneous competitions, would provide an excellent use case for the system.

The VSAR project (a UK Government Technology Strategy Board funded consortium of the BBC, Technium CAST, BAE Systems and the National Physical Laboratory) was constituted to explore the possibilities of this new medium for both entertainment and security applications. The goal of the project is to develop a suite of software tools that will allow a third party to create, deliver and render live, 3D interactive models generated from real-life. The project aims to complete in April 2011. This paper concentrates on the entertainment and domestic use of the system, although most of the functionality is identical for potential security applications.

2 Home delivery of 3D models

A home today might contain a number of hardware platforms that could display interactive 3D models: mobile phones, games consoles and PCs are all capable of rendering 3D graphics (and in the near future, set-top digital TV decoder boxes may also be able to do so). This platform profusion can be both a blessing and a curse, but we have concentrated our current development on a PC-based system since it is likely to be the most common piece of suitable domestic equipment and offers a fairly standard interface to the application developer via a web interface.

Since the BBC is a mass market media provider, it is very important that our web services and applications are as easy as possible for our audiences to access, particularly where non-technical people are asked to try something new. With this in mind, and with the recent developments in ActionScript-based open source 3D engines, such as Away3D [3] and Papervision3D [4], we chose to implement a prototype VSAR renderer in ActionScript 3 for delivery via Adobe Flash [5] so that it could be viewed in a standard web browser.

An initial concept Flash-based demonstrator, showing how a 3D Wimbledon service might look was created using Away3D, a simple keyhole markup language-zipped (KMZ) format model of the Wimbledon environment [6] and a number of short clips of tennis matches taken from a vantage point at the end of the court.

The video clips were converted from an MPEG transport stream into Flash Video (FLV) format [7] by FFMPEG [8] so that they could be decoded in Flash and rendered onto billboards in the 3D model. These billboards, and associated viewing cameras were placed ‘by hand’ in the model so that the perspective and scale of the videos matched the background model geometry when viewed from the location of the camera. A few simple controls allowing the user to ‘fly’ between the camera positions and view the action on different courts were added to allow interactive navigation. This simple demonstrator (Figs. 1 and 2) proved surprisingly effective in giving the illusion of ‘flying into’ the scene to view the live action and offered exciting possibilities for navigating between the matches.

Although this demonstrated the concept of inserting videos into a 3D model using existing web technology there were a number of limitations that needed to be overcome.

First, if the real camera altered its pose, there was no way to reflect that in the pose of the virtual camera, resulting in the foreground video ‘slipping’ against the background model. This problem is normally dealt with using a camera tracking system [9] or a pan-tilt sensor head. These technologies work effectively, but require specific hardware and access to the camera telemetry data. If one is receiving third party feeds, or using archive material, this data will not be available, so some method of tracking camera location purely from analysing the video is required. The method should provide not just relative motion from frame-to-frame, but absolute position and angles in the reference frame of the 3D model.

Secondly, the user was constrained to viewing the action from a specific viewpoint. By measuring the location of the

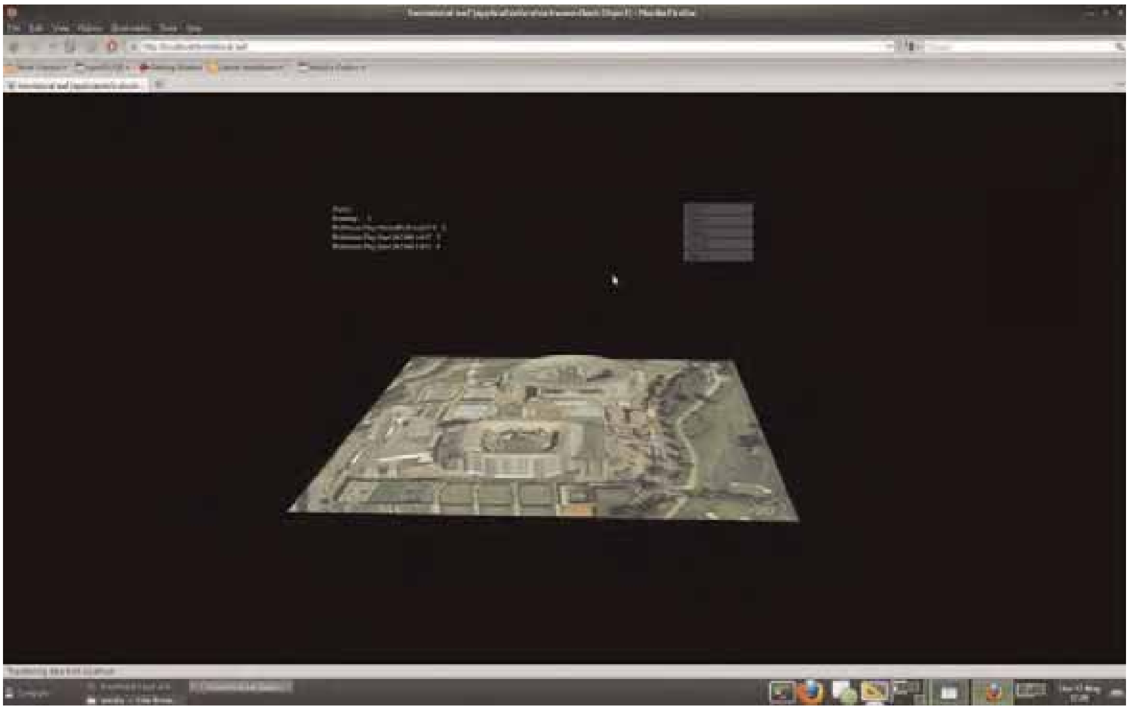


Figure 1 *Prototype VSAR web interface, running as a Flash application in a standard web browser, showing 3D model of Wimbledon Lawn Tennis Club*

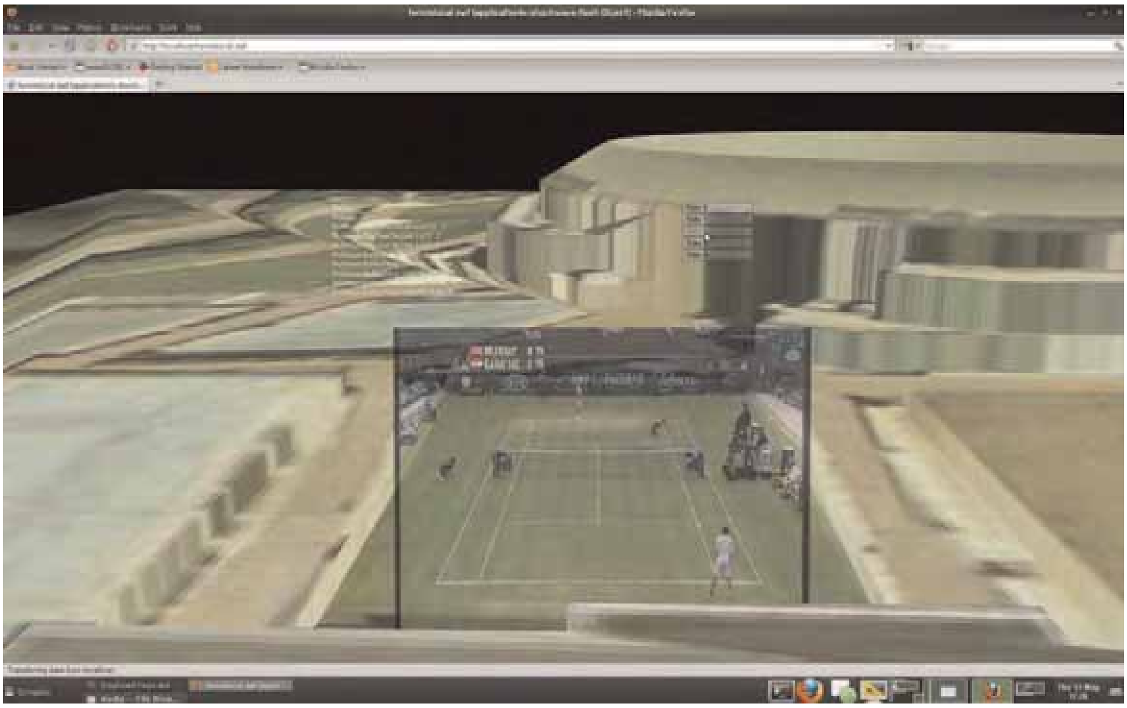


Figure 2 *Prototype VSAR web interface, showing a video billboard overlaid on the 3D Wimbledon model*

players in the scene and placing a billboard of their texture extracted from the video at their actual location in the model a wider range of viewing positions for the user can be supported. This latter technique was first developed in the Prometheus project [10] and is now used for rendering reconstructions of football and rugby games [11]. A more advanced technique could be used, that builds true 3D models of the foreground action (rather than flat billboards), using images from 4–6 cameras from different viewpoints, as described in Grau *et al.* [12]. However, for the application areas considered here, there are unlikely to be multiple cameras viewing the same foreground region, so single-camera methods are used. In previous sports-based applications, locations of the players are determined by generating a key signal from the green grass background and estimating their 3D location using the assumption that the lowest point in the key (usually the feet) is in contact with the ground plane. This keying method was not suitable in this application, as the scene background could be any arbitrary environment. So alternative methods of detecting people and objects in images needed to be developed.

Finally, all this extra data, camera pose and field of view, and the 3D location of objects in the video, needs to be communicated to the renderer. The solutions to these problems are discussed below.

3 Real-time camera pose measurement from image analysis

In any 'mixed-reality' video system, where rendered elements are being inserted into a real 3D scene, it is necessary for the virtual (rendering) camera to match the position, orientation and field of view of the real camera (referred to as the camera pose) for the view, scale and perspective of the virtual elements to be rendered correctly. For a non-static camera, this information needs to be measured for every frame. VSAR could use camera telemetry data from physical sensors but provides a specific implementation of a video-based camera tracker, allowing use on arbitrary cameras without special mountings or return channels for data. Previous work, targeted at virtual graphics for sports such as football and rugby, used the presence of pitch lines at known positions to compute the camera pose [13]. In the more general situation that VSAR considers, it was necessary to extend this approach to use arbitrary image features, such as corners of buildings and areas of rich image detail.

The camera tracker developed in VSAR [14] requires a one-off manual identification of features in the image that have known positions in the world, such as lines on an athletics track, or corners of a stadium. This allows the initial camera pose to be computed. A GUI has been developed that shows a simplified 3D model of the scene incorporating these features, and allows an operator to select a 3D feature and then identify its corresponding location in the camera image. The camera is then panned round the scene to cover the entire area of interest, and suitable features for tracking (such as corners) are automatically identified and stored. These features are allocated 3D coordinates consistent with their location in the first image in which they are seen. The depth assigned to each feature is arbitrary, as the tracker assumes the camera is rotating around a fixed point. Additional points or lines in known positions can be added during this process to improve the estimate of the camera pose and the locations assigned to the features.

Once a set of features have been learned, the tracker can initialise itself by matching stored features against those in the current camera view. It then tracks the camera motion from frame-to-frame, using the well-known Kanade–Lucas–Tomasi (KLT) tracker to match features. The matching process uses a combination of image patches around features as they were first seen (to avoid long-term drift) and image patches around features as seen in a recent image (to improve robustness against effects such as motion blur and lighting change). The tracking algorithm runs at full video rate on a single processor of a modern PC. Further details may be found in Dawes *et al*. [14].

4 Measuring 3D object positions from a single camera

The intersection of two vectors in a volume can provide an unambiguous 3D coordinate, but it would be inconvenient to require two cameras to cover every location from where 3D coordinates might be required and it would be difficult to match objects from two images captured at different angles. Therefore, VSAR makes the assumption that the object being detected is standing on the ground. The intersection of a single vector from the camera at the ground plane therefore gives a pseudo-3D coordinate. The problem is then reduced to actually detecting the position of objects in a video image, and determining which part of their geometry is in contact with the ground. Since VSAR could conceivably be deployed in any environment, the ‘green grass’ keying system used in previous sports-based applications could not be relied upon. Therefore, a previously developed algorithm for ‘difference keying’, (which compares the appearance of successive video images to generate a key signal,) was used. Difference keying is effective in environments where lighting can be controlled (such as building interiors) but is less reliable in outdoor environments where clouds occluding the sun can cause rapid lighting intensity and colour changes. To provide a more reliable system for detecting the location of objects in a scene, we investigated an implementation of the technique of histogram of orientated gradients (HOG) [15, 16] using the OpenCV [17] computer vision library, and trained it on standing human shapes using the INRIA dataset [18]. In many applications we know that the objects of interest will be people, but the algorithm can be trained on the shapes of other objects if required. This algorithm proved to be very reliable at locating people in the image for a wide variety of image contrasts and brightnesses, although it was processor intensive, taking approximately 1.5 s on a PC with two dual core Opteron 2220 2.8 GHz processors to process a single 720 × 576 pixel image. The bulk of this processing time was owing to the requirement to ‘scan’ the 64 × 128 pixel HOG filter over the image to locate the actual position of the people. This time was reduced by using the difference keyer to detect areas of motion and limiting the HOG search to those regions. This reduced the processing time to ∼0.2 s or 5 fps.

Recall that it is necessary to detect the intersection of the line-of-sight to the lowest point of the detected object with the ground plane to calculate a 3D coordinate, therefore we need to measure the location of the detected people’s feet. This was done by performing a horizontal Sobel filter on the video image, and then thresholding the regions where human shapes had been detected. This created a binary image which highlighted the strong horizontal edge of the person’s lower leg, the spatially lowest positive pixel value within the lower quarter of the detected region was then assumed to be the position of the foot. Once the closest pixel to a person’s foot (or base of object) is known, a vector is projected from the camera image plane through the pixel. Since the camera position and orientation is known, the intersection of this vector with the ground plane gives the location of the object on the ground.

5 VSAR data encoding and transmission

The VSAR web-based system uses Flash as a front-end renderer and FLV or XML as the data delivery formats.

FLV provides encapsulation for a variety of video and audio formats, but also allows for the addition of frame-by-frame metadata. This is typically used to ‘cue’ actions in a Flash presentation but also allows the addition of any number of key:value string pairs (provided the total size of the data frame is less than 0 × FFFFFF bytes). These key:value string pairs are used by VSAR to carry data about the camera and the positions, sizes and types of objects in the 3D scene. This data is extracted by the renderer and used to generate the presented scene.

VSAR uses the × 264 open source encoding library [19] to encode its video into H.264 (AVC) format [20] for distribution. H.264 is widely used in delivering video for domestic consumption over the web owing to its flexibility in low bit rate applications. It is also increasingly used in CCTV surveillance applications owing to the significant image quality improvements that it offers over other compression formats. It is one of the video formats supported in FLV.

Audio is encoded using the freeware advanced audio coder (FAAC) [21] by the VSAR system. FAAC encodes audio using the advanced audio codec (AAC) [22], a format widely supported in consumer devices owing to its perceived superiority to MP3 for low bit rates. AAC is also supported by FLV.

VSAR provides an interface for inserting data from other systems, such as GPS or audio localisation devices into the Flash stream. This data could be extracted by the renderer, perhaps to provide location tracks of objects – or audio ‘hotspots’ showing where specific or loud noises have occurred.

For applications that require only the metadata information (such as an application on a hand-held device showing a map and the position of detected objects relative to a user) VSAR can provide an XML data feed.

6 System overview

A complete VSAR system consists of a number of components for capturing and presenting a 3D scene to a user. Video, audio and other data must be captured, analysed, encoded for IP transmission, delivered, decoded and displayed. Fig. 3 shows the components of a basic VSAR installation.

In this example, the video and audio data from the cameras and microphones are being encoded into FLV format by the processing PCs, along with any extra metadata generated from audio or video analysis. The FLV streams are made available by VSAR via a TCP or UDP network connection. This provides considerable flexibility of use, since the output streams can be routed to other devices for further analysis (and metadata insertion), or to a content distribution network for widespread dissemination, or directly to an instance of a VSAR renderer for display.

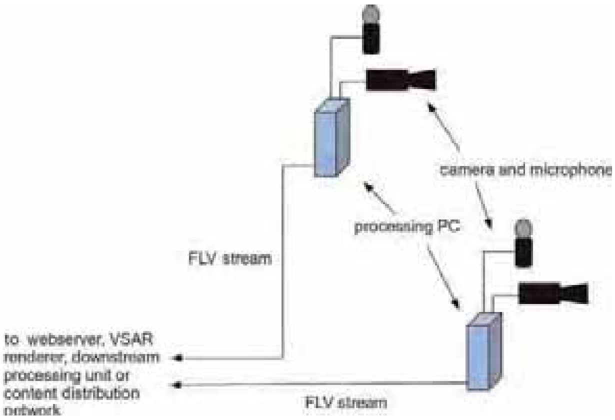

camera and microphone
processing PC
FLV stream
to webserver, VSAR renderer, downstream processing unit or content distribution network
FLV stream

Figure 3 *Basic VSAR system, setup for a two camera shoot*

More complex systems are possible, including those using hundreds of cameras, by taking advantage of the VSAR Governor. The VSAR Governor is a geospatial database and control system that supports the live addition and removal of cameras and other data sources, selection of camera feeds by required geographical area of coverage and automatic tracking of objects of interest.

An alternative, high quality renderer for high-value installations has also been developed based on the OpenSceneGraph graphics tool-kit [23].

7 Results

Fig. 4 shows the output from the HOG person detector and foot position finder. The green boxes show where the output from the HOG filter was above a threshold determined empirically. Note that the filter has been trained on standing human shapes only. The horizontal red line shows the height of the lowest positive pixel value from the

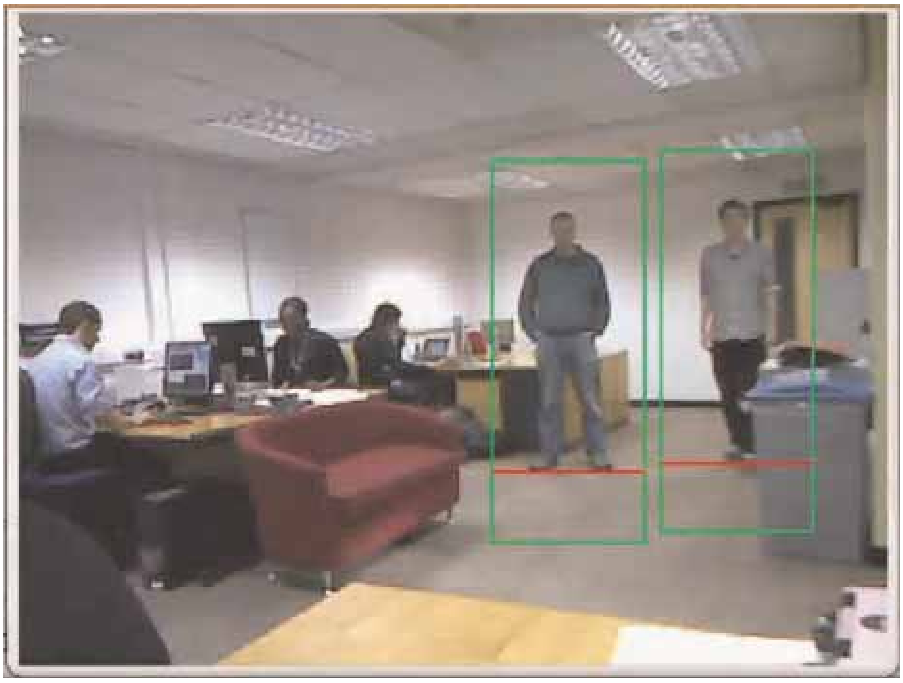


Figure 4 *Output from the HOG person-detection algorithm and Sobel filter foot-finder*

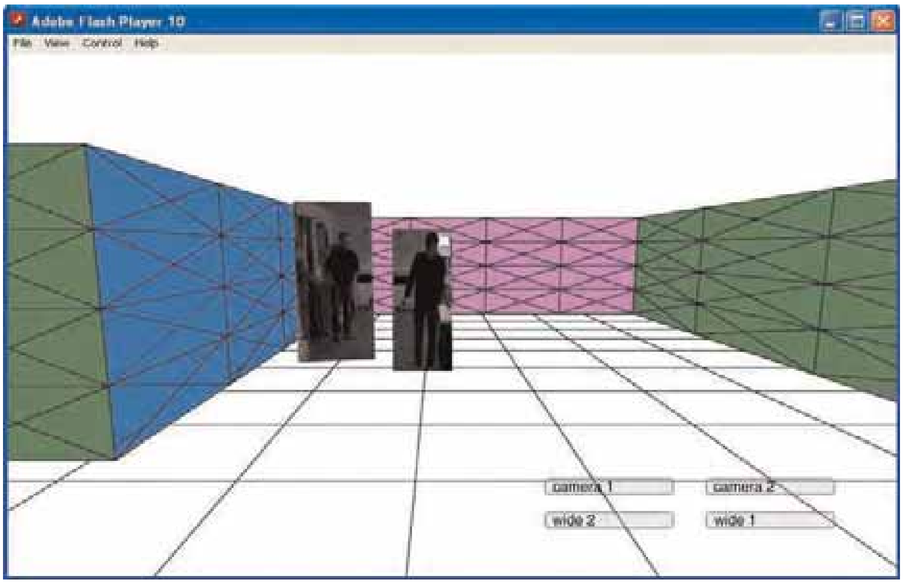


Figure 5 *Video billboards of detected people, inserted into 3D model of their environment*

horizontal Sobel filter, in this case corresponding to the position of the foot in the image.

Fig. 5 shows two video billboards inserted into a 3D model of the office in which they were filmed. The area of the video texture on the billboards is determined by the output from the HOG detector. The billboards are inserted into the model at the locations calculated from the calibrated position of the camera and the output from the Sobel filter-based foot-finder. The viewing position of the virtual camera corresponds to the actual position of the real camera in the office.

Fig. 6 shows the view from a virtual camera in a model of an office. The position of the virtual camera matches that of the real camera, and the vertices of the 3D model match those of the real office. The camera location was calibrated by matching vertices in the 3D model with features visible in the camera image. A video billboard from the footage of the real camera has been added to the scene. The billboard is scaled to exactly match the boundaries of the real camera viewing frustum, although the virtual camera can be zoomed independently to provide a wider-than-real-life view of the scene. Note, the close correspondence between the walls in the model and the walls in the video. The pose of the real camera is calculated for each frame by the VSAR camera tracker and is used to update the pose of the virtual camera.

Although it is necessary to view the scene with a virtual camera with the same pose as the real camera in order for the inserted video to correspond to the 3D model geometry,

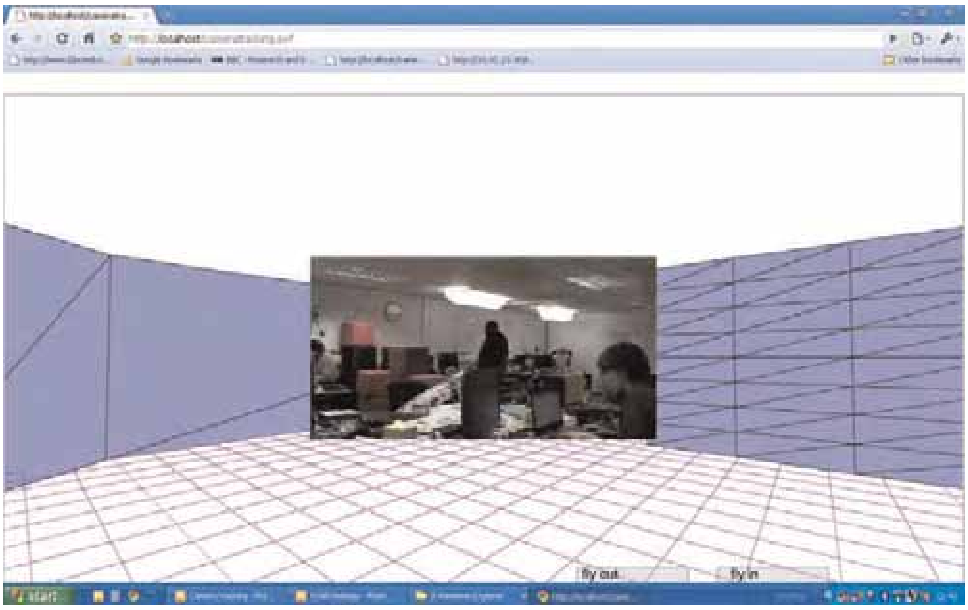


Figure 6 *Video billboard inserted into 3D model corresponding to actual geometry of the room in which it was filmed*

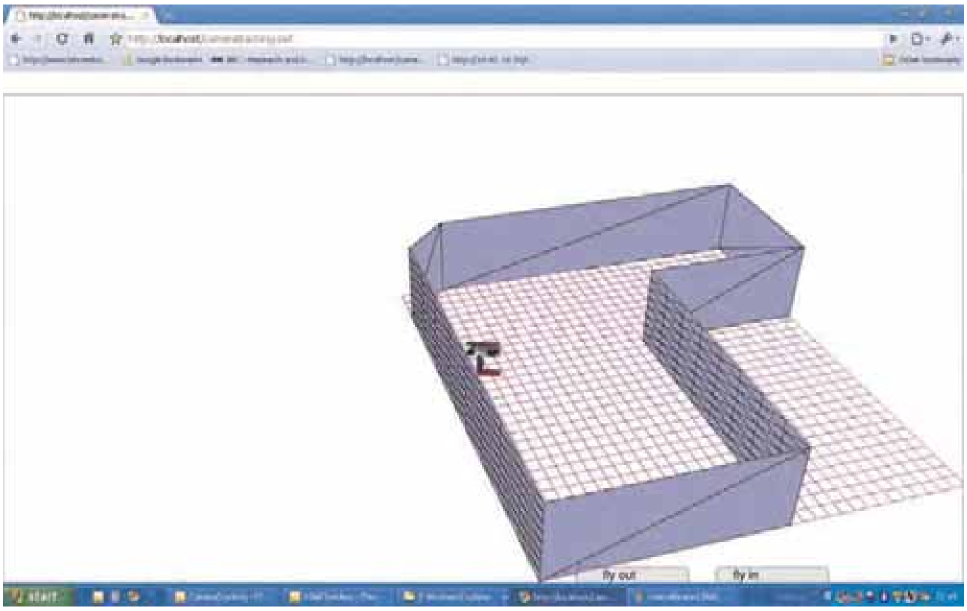


Figure 7 *Alternative view of model show in Fig. 6*
Position and orientation of virtual camera are marked by axes near centre of the image

it is, of course, possible to view a 3D model from any position. Fig. 7 shows an alternative view of the model and billboard in Fig. 6. This kind of navigation is effective at providing an understanding of the wider geographical environment.

8 Conclusions

We have shown that merging live 2D video and pre-generated 3D models provides an interesting new way of representing a live scene, in a way that does not need multiple overlapping cameras to create a full 3D foreground model. We have also demonstrated that it is possible to deliver these models directly to a standard web browser using existing technology.

Further work will include looking at the use of pre-rendered video fly-throughs (giving higher quality backgrounds with less reliance on the power of the user's PC, but with the limitation of pre-selected viewpoints and camera paths), and looking at ways of increasing the realism of video textures for objects that are not well-represented as billboards (such as a dense crowd).

This technology is also being explored in the context of supporting security personnel who have to view large arrays of CCTV cameras dispersed over a large geographic area. In this case the CCTV camera views would be overlaid onto a 3D model of the facility. The purpose of the technology would be to improve the ability of security personnel to identify threats or problems, track dynamic events over large areas, coordinate more effectively with personnel on the ground, and improve their overall ability to assess, make sense of evolving events and act on their assessments. Further research is being undertaken by BAE Systems to establish if the combination of video and 3D models created by VSAR is effective in improving these tasks or whether the system is merely 'eye candy', making it harder for the security operators to do their jobs. To do this, appropriate measures of effectiveness for the various aspects of the task must be identified. For example, measures might include: speed of noticing an unusual 'target' (and the trade off with potential for false alarms), ability of an operator to track a moving target over more than one camera 'eye-span' without losing the target, and the efficiency of providing directions to personnel 'on the ground' via radio or mobile devices. This sort of evaluation will presumably also be helpful for understanding the value to viewers of this technology for entertainment purposes. However, the measures of 'enjoyment', 'engagement' or 'fun' are less obvious. Measurement techniques will likely include psychophysiological measures [24], opinion surveys, and other 'human factors' methods [25], many of which have been designed to evaluate virtual reality and gaming experiences [26]. The current project will attempt to understand some of these proposed user enhancements.

9 References

[1] Bamzooki: http://www.bbc.co.uk/cbbc/bamzooki/

[2] ITN Virtual Reality Studio: http://corporate.itn.co.uk/itn-news/itn-studios.aspx

[3] Away3D: http://www.away3d.com/

[4] Papervision3D: http://blog.papervision3d.org/

[5] Adobe Flash: http://en.wikipedia.org/wiki/Adobe_Flash

[6] Wimbledon Model, by Kevin Girard: http://sketchup.google.com/3dwarehouse/details?mid aacd13d74d14056eb9060b18736fffd1

[7] Flash Video (FLV) Format Specification: http://www.adobe.com/devnet/flv/pdf/video_file_format_spec_v10.pdf

[8] FFMPEG: http://www.ffmpeg.org/

[9] THOMAS G.A., JIN J., NIBLETT T., URQUHART C.: 'A versatile camera position measurement system for virtual reality TV production'. IEE Conf. (Publication No. 447), IBC, Amsterdam, September 1997, pp. 284–289

[10] PRICE M., CHANDARIA J., GRAU O., *ET AL*.: 'Real-time production and delivery of 3D media'. Proc. Int. Broadcasting Convention, Amsterdam, Netherlands, September 2002, available as BBC R&D White Paper 045, http://www.bbc.co.uk/rd/publications/whitepaper045.shtml

[11] Red Bee Media Ltd: The Piero sports graphics system, www.redbeemedia.co.uk/piero, www.bbc.co.uk/rd/projects/virtual/piero/

[12] GRAU O., HILTON A., KILNER J., *ET AL*.: 'A free-viewpoint video system for visualisation of sport scenes'. Proc. IBC 2006, Amsterdam, NL, 7–11 September 2006, available as BBC R&D White Paper 142, http://www.bbc.co.uk/rd/publications/whitepaper142.shtml

[13] THOMAS G.A.: 'Real-Time Camera Tracking using Sports Pitch Markings', *J. Real Time Image Process.*, 2007, **2**, (2–3), pp. 117–132, available as BBC R&D White Paper 168. http://www.bbc.co.uk/rd/publications/whitepaper168.shtml

[14] DAWES R., CHANDARIA J., THOMAS G.A.: 'Image-based camera tracking for athletics'. IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting (BMSB2009), Bilbao, 13–15 May 2009, available as BBC R&D White paper 181 http://www.bbc.co.uk/rd/publications/whitepaper181.shtml

[15] DALAL N., TRIGGS B.: 'Histograms of oriented gradients for human detection', *CVPR*, 2005, **1**, pp. 886–893

[16] ZHU Q., AVIDAN S., YEH M C., CHENH K T.: 'Fast human detection using a cascade of histograms of oriented gradients', *IEEE Comput. Vis. and Pattern Recognit.*, 2006, **2**, pp. 1491–1498

[17] OpenCV: http://opencv.willowgarage.com/wiki/

[18] INRIA dataset: http://pascal.inrialpes.fr/data/human/

[19] X264 video coder: http://www.videolan.org/developers/x264.html

[20] H264 specification: 'ISO/IEC 14496-10:2008 Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding', 2008

[21] FAAC audio coder: http://www.audiocoding.com/

[22] AAC specification: 'ISO/IEC 13818-7:2003 Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)', 2003

[23] OpenSceneGraph: http://www.openscenegraph.org/projects/osg

[24] DRACHEN A., NACKE L.E., YANNAKAKIS G., PEDERSEN A.L.: 'Correlation between heart rate, electrodermal activity and player experience in first-person shooter games', 2010, http://hci.usask.ca/uploads/176-DrachenNackeetal_Correlations.pdf (accessed May 10, 2010)

[25] JACKO J.A., SEARS A. (EDS.): 'The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications' (Lawrence Earlbaum Associates Inc, New York, NY, 2003)

[26] NACKE L., DRACHEN A., GOEBEL S.: 'Methods for evaluating gameplay experience in a serious gaming context', *Int. J. Comput. Sci. Sport*, 2010, **9**, (2010), http://hci.usask.ca/publications/view.php?id 174 (accessed May 10, 2010)

Selected content from IBC 2010

ISSN 2041-5923

Robust and low-complexity detection technique for DVB-T/H receivers in fast fading channels

L. Zhang Z. Hong L. Thibault

Communications Research Centre, 3701 Carling Avenue, Ottawa, ON K2H 852, Canada
E-mail: liang.lu-zhang@crc.ca

Abstract: Digital video broadcasting terrestrial/handheld (DVB-T/H) systems are deployed worldwide to deliver multimedia broadcasting to static and mobile receivers respectively. It is desirable to deploy the DVB systems in 8K mode, which provides the largest transmitter separation in a single-frequency network (SFN) configuration. It is also desirable to use 64 QAM with higher channel coding rates for data transmission to achieve the best spectrum efficiency. However, since these systems employ orthogonal frequency-division multiplexing (OFDM), fast-moving DVB-T/H receivers suffer significant performance degradation owing to the inter-carrier interference (ICI). Currently, the ICI is the major impairment which prevents good mobile reception for DVB services deployed with 8K mode and 64 QAM modulation.

In this paper, the authors describe and low-complexity decision-directed detection technique which significantly improves the performance of fast-moving DVB receivers. Simulation results show that, for DVB-T transmission at 800 MHz with 8K mode, 64 QAM modulation and rate-1/2 convolutional code (CC), mobile receivers equipped with the proposed technique and a single antenna can achieve the required quality of service (QoS) for speed as high as 180 km/h. This guarantees good reception for most car receivers (or portable receivers in cars). For a DVB-T receiver with two receive antennas, even with a weaker rate-2/3 CC, the proposed technique can provide robust performance for receiver speed up to 200 km/h.

1 Introduction

Digital video broadcasting terrestrial/handheld (DVB-T/H) systems [1, 2] have been deployed worldwide to deliver high data rate television and multimedia broadcasting services to stationary and mobile receivers. Orthogonal frequency division multiplexing (OFDM) is employed in DVB-T/H systems because it can effectively overcome the multipath fading of mobile wireless channels. In an OFDM system, a wideband frequency-selective channel is divided evenly into a group of narrowband nearly flat-fading subchannels. Each subchannel carries a modulated data symbol, which could be QPSK, 16 QAM or 64 QAM symbol. Known pilot symbols are inserted in a number of specific subchannels to help receivers in estimating the subchannel gains. Most current DVB-T/H receivers perform conventional coherent detection (CCD), where the data symbol on the k^{th} subcarrier is recovered by a one-tap equaliser as

$$Y_{eq}(k) = \frac{Y(k)}{H_{est}(k)} \tag{1}$$

where $Y(k)$ and $H_{est}(k)$ are the received symbol and the estimate of the multiplicative gain of the k^{th} subchannel. The subchannel gains are estimated from the received in-band pilots.

While providing good performance in time-invariant wireless channels, CCD-based DVB receivers are very vulnerable in time-varying channels experienced by moving receivers. The channel variation results in Doppler spread in the received signal, which causes signal power leakage from one subchannel into its neighbouring subchannels. This power leakage is called inter-carrier interference (ICI).

For one subchannel, the ICI from all the neighbouring subchannels is another source of noise and could significantly degrade the detection performance. Since the ICI is generated from the transmitted signal itself, increasing transmission power will not solve this problem. This results in an error floor in the receiver performance. This error floor becomes more severe for DVB receivers moving at higher vehicle speeds. There exists a speed limit above which CCD-based receivers can no longer achieve satisfactory reception, no matter how much power is transmitted. For DVB transmissions, this speed limit depends on the sub-carrier spacing (i.e. the transmission mode), the spectrum efficiency (modulation and channel coding schemes), the RF carrier frequency, and the type of mobile environment. Currently, ICI is the major impairment which prevents good mobile reception for DVB services deployed with 8K mode and 64 QAM modulation.

To achieve better mobile reception for DVB-T receivers, a novel low-complexity COFDM detection technique (hereafter referred as CRC DVBDetect) was developed at the Communications Research Centre (CRC), Canada, which can effectively remove a significant part of the ICI in the received signal. It is shown by simulation that, with CRC DVBDetect scheme, a DVB-T receiver with a single antenna operating in the 8K mode with 64 QAM and rate-1/2 CC can achieve satisfactory performance when moving at speeds up to 180 km/h with RF transmission at 800 MHz. For a dual antenna DVB-T receiver using maximum ratio combining (MRC) and CRC DVBDetect, good reception is achieved for speeds up to 200 km/h with a weaker rate-2/3 CC.

Although DVB-T receiver is used in our investigation, the CRC DVBDetect advanced detection technology is directly applicable to DVB-H receivers. In this case, it is not necessary to activate the MPE-FEC for the DVB-H receivers to obtain good performance. This can improve the bandwidth efficiency and reduce the end-to-end service delay caused by the huge interleaver of the MPE-FEC.

The paper is structured as follows. In the next Section, we briefly describe the DVB-T system and the problem associated with the use of 8K mode and 64 QAM. Then the CRC DVBDetect technique is introduced, followed by the simulation performance of DVB-T receivers with the proposed technique in fast fading channels, with either one or two receive antennas. Conclusions are drawn in the last Section of the paper.

2 Problem of mobile reception in DVB-T system

As defined in the ETSI DVB standards [1, 2], a DVB-T/H signal is made of a sequence of OFDM symbols, where each OFDM symbol consists of a guard interval (GI) followed by the 'useful' signal part. Service data is carried in data subchannels with one of the modulation schemes, i.e. QPSK, 16 QAM or 64 QAM. A group of evenly distributed subchannels are allocated to carry the in-band pilots. Two transmission modes are defined in [1] for DVB-T system, each having a different set of transmission parameters. In Table 1, we list the parameters for the two transmission modes of DVB-T systems with 6 MHz bandwidth.

Designed to 'absorb' the inter-symbol interference created by multipath, the GI also allows the simultaneous transmission of the same DVB signal by many transmitters using the same frequency in so-called single-frequency networks (SFNs). SFNs are very attractive to broadcasters since they are spectrum efficient and provide a form of transmit diversity to the receivers. In a SFN, the duration of the guard interval determines the maximum distance between transmitters. It is shown that the GI's for 8K mode is four times as long as the GI's in 2K mode. Therefore, the maximum transmitter separation for 8K mode is four times larger than that for 2K mode. The 8K mode is more economic for DVB-T SFN deployment in terms of the infrastructure cost, i.e. it takes much less transmitters for 8K mode to cover the same service area than 2K mode.

Performance of DVB receivers at high vehicle speeds is limited by the ICI generated by the Doppler spread. The level of channel variation can be characterised by the normalised Doppler spread f_dT_u, where f_d is the maximum Doppler shift, T_u is the useful OFDM symbol duration and $1/T_u$ is the subcarrier spacing. The higher the f_dT_u value, the worse the performance. The maximum Doppler shift, f_d, is directly related to the vehicle speed, v, and the carrier frequency, f_c, as, $f_d \quad f_c \cdot v/c$, where c is the light speed. As shown in Table 1, the 8K mode has four times the useful symbol duration of the 2K mode, 1195 μs against 299 μs. For the same maximum Doppler shift f_d,

Table 1 Parameters of the two transmission modes of 6 MHz DVB-T system

Parameters		Mode	
		2K	8K
K	number of subcarriers	1705	6817
N	FFT size (points)	2048	8192
T_u	useful symbol duration (μs)	299	1195
T_G	guard interval duration (μs)	[75 37 19 9]	[299 149 75 37]
B	total bandwidth	5.71 MHz	
$1/T_u$	subcarrier spacing (kHz)	3.384	0.837

i.e. the same vehicle speed, the 8K mode has a normalised Doppler spread ($f_d T_u$) value four times that of the 2K mode. Therefore, the 8K mode suffers significantly more performance degradation in mobile reception.

The effect of ICI also depends on the spectrum efficiency. While 64 QAM provides the best bandwidth usage, it is more susceptible to ICI, as compared to QPSK and 16 QAM. Selecting a higher rate convolutional code can also provide better spectrum efficiency, which unfortunately also proved to be more sensitive to ICI in fast fading channels.

To achieve both good spectrum efficiency and low infrastructure cost, it is necessary to use 8K mode, 64 QAM modulation and higher rate convolutional code. However, it is shown in Gerard [3] and Ralf *et al.* [4] that for a DVB-T channel at around 640 MHz, with rate-1/2 convolutional code (CC), using the 8K mode and 64 QAM, mobile receiver speed is limited to 50 km/h. This suggests that most moving vehicle receivers cannot obtain good quality DVB-T services.

In the next Section, we describe an advanced COFDM signal detection technique for DVB-T/H signals which makes high data-rate DVB-T services with 8K mode and 64 QAM available to fast-moving receivers.

3 Advanced COFDM detection technique for mobile reception

The ETSI standards [1, 2] only defines the DVB broadcasting system on the transmitter side, putting no constraint on receiver design. Therefore, advanced signal processing techniques can be designed and incorporated in DVB receivers for better performance.

It is mentioned earlier that the speed limit is imposed by ICI, which is a consequence of the Doppler spread. In a conventional DVB receiver, ICI is considered as additive noise source. However, contrary to thermal noise, ICI does have a systematic structure. With the knowledge of the transmitted data and the wireless channel condition, it is possible for the receiver to generate a replica of the ICI on each subchannel and to cancel it from the received signal. This approach is called ICI cancellation and can significantly lower the error floor, and therefore improve the service availability for moving receivers.

A simplified block diagram of CRC DVBDetect scheme is shown in Fig. 1. For the receiver to re-construct the ICI, it is necessary to know the transmitted data and the wireless channel condition, neither of which is available to the receiver. Therefore, an iterative process is proposed. As shown in Fig. 1, in the first iteration, the receiver carries out CCD, i.e. performing the pilot-aided channel estimation, followed by the standard signal detection processes (one-tap equalisation, demodulation, deinterleaving and Viterbi decoding) to make a tentative decision on the transmitted bit sequence. Estimates of the subchannel data symbols are re-constructed from this tentative decision bit sequence, by convolutional encoding, time-domain interleaving and QPSK/QAM modulation. These estimates of the subchannel data symbols are hereafter referred to as decision-feedbacks.

The core of CRC DVBDetect is an efficient technique to accurately estimate the channel characteristics necessary to reconstruct the ICI, i.e. the ‘decision-directed channel estimation & ICI cancellation’ block in Fig. 1. With relatively low complexity, this technique can provide a very accurate channel estimation. Specific implementation details of this technique will not be further discussed here since it is in the process of being patented.

With the decision-feedbacks $\tilde{X}$ and the channel estimate $\tilde{\tilde{F}}$, the receiver can now re-construct an estimate of the ICI and subtract it from the received data symbols to generate an ICI-reduced symbol sequence, Y_{ICIred}. This sequence is then passed to the standard signal detection blocks to generate more reliable decisions. Additional iterations can

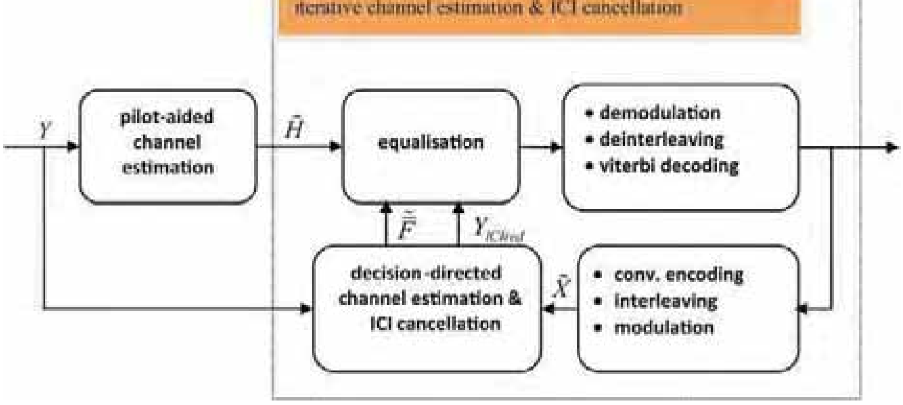


Figure 1 *Simplified block diagram of CRC DVBDetect technology*

be carried out based on this more reliable decision to obtain further performance improvement.

It is shown in [3] that receive antenna diversity is another technique to improve OFDM receiver performance in fast fading channels. CRC DVBDetect can also be combined with receive antenna diversity in DVB receivers to provide further performance improvement.

4 Simulation

To demonstrate the effectiveness of CRC DVBDetect, the performance of a DVB-T receiver with and without CRC DVBDetect was evaluated by computer simulations, assuming DVB-8K mode with 64 QAM modulation. A BER of $2 \cdot 10^{\ 4}$ at the output of the Viterbi decoder is used as the criteria for good quality of service (QoS), which results in 'virtually error free' performance at the output of the outer Reed–Solomon decoder, Yannick [5]. The wireless channel is generated according to the typical urban (TU) channel model defined in the COST 207 report [6], which has an RMS delay spread (DS) of 1 μs and a maximum delay of 7 μs. Each BER measure is an average over 10 distinct TU channel realisations. The DVB channel is assumed to have a bandwidth of 6 MHz at a carrier frequency of 800 MHz. This provides a worst case scenario since 800 MHz is almost the highest UHF TV frequency and the 6 MHz bandwidth has the narrowest subchannel bandwidth, both contributing to the highest ICI impact.

4.1 Fading channel model

The TU channel is modelled as the sum of $M \quad 100$ paths with equal attenuation. Each path is associated with a phase shift, a Doppler frequency shift and a path delay. The random phase shift is a random variable uniformly distributed in $[0,\ 2\pi)$. A classical 'U' shaped Doppler spectrum is assumed for each path and is defined by

$$p(f) = \frac{1}{\pi\sqrt{f_d^2 - f^2}} \tag{2}$$

when $|f| \leq f_d$ and is zero for other values of f. f_d is the maximum Doppler shift.

The delay value of each path is a random variable taken from a exponential distribution defined in [6] as

$$p(t) = \frac{1}{\sigma_d} e^{\ t/\sigma_d} \tag{3}$$

when $t \geq 0$ and is zero for negative values of t and for values higher than σ_{max}. σ_d is the RMS delay spread of the channel. For TU channels, $\sigma_{max} \quad 7$ μs, and $\sigma_d \quad 1$ μs. The exponential power delay profile (PDP) defined in [6] is thus realised by generating the delay values from this distribution and assigning equal power to each path.

4.2 Simulation results

In Fig. 2, the performances of DVB-T receivers with different detection techniques and number of receive antennas are presented in terms of the SNR requirements to achieve the desired QoS (a BER of $2 \cdot 10^{\ 4}$ at the output of the Viterbi decoder) against vehicle speed. Simulation results are presented for the following six scenarios:

- CCD, r 1/2: conventional detection with a single antenna, rate-1/2 CC

- CRC DVBDetect, r 1/2: CRC DVBDetect technique with a single antenna, rate-1/2 CC

- CCD, r 2/3: conventional detection with a single antenna, rate-2/3 CC

- CCD, r 2/3, maximum ratio combining (MRC): conventional detection with dual antenna MRC, rate-2/3 CC

- CRC DVBDetect, r 2/3: CRC DVBDetect technique with a single antenna, rate-2/3 CC

- CRC DVBDetect, r 2/3, MRC: CRC DVBDetect technique with dual antenna MRC, rate-2/3 CC

Fig. 2 shows that when a rate-1/2 CC is used, a conventional receiver with a single antenna requires an E_b/N_0 of 14.6 dB to achieve the desired QoS for very slowly moving receivers (1.2 km/h). The SNR requirement increases with receiver speed. For speed of 90 km/h or higher, the receiver can no longer achieve satisfactory performance. A receiver with CRC DVBDetect and one receive antenna requires an SNR of 14 dB for slowly moving receivers. The SNR requirement increases much slower with receiver speed as compared to the conventional

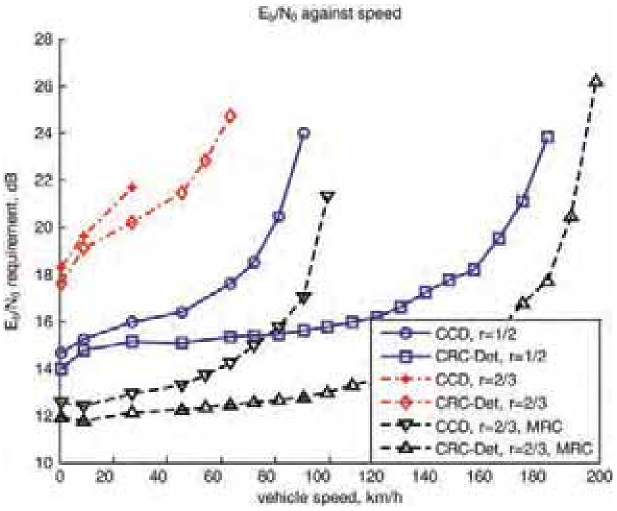


Figure 2 *E_b/N_0 requirement against mobile speed to achieve a BER of $2 \cdot 10^{\ 4}$ at Viterbi decoder output for different reception scenarios, 8K mode, 64 QAM, RF@ 800 MHz, 6 MHz bandwidth, TU channel*

receiver. An E_b/N_0 of 16.3 dB is sufficient to achieve the desired QoS for speed up to 120 km/h and 18 dB for receiver moving at 150 km/h. The speed limit is extended to 180 km/h, which meets the requirements for car receivers or portable receivers used in cars. The speed performance of CRC DVBDetect with dual antenna and rate-1/2 CC was not evaluated but can be expected to exceed 300 km/h and be suitable for high speed trains.

The use of rate-2/3 CC has a better spectral efficiency but significantly worse performance than rate-1/2 CC. It is shown in Fig. 2 that with 64 QAM, rate-2/3 CC and a single receive antenna, conventional DVB-T receivers require an E_b/N_0 of 18 dB to achieve the desired QoS at very slow speed. CCD-based receivers cannot achieve good QoS at speeds faster than 30 km/h. For receivers with CRC DVBDetect, the speed limit is extended to 65 km/h, which is unsatisfactory for car receivers.

We propose to use antenna diversity to achieve further improvement in mobile reception of DVB-T. In Fig. 2, we plotted the performance curves for a dual antenna receiver with MRC. For rate-2/3 CC, it is shown that when conventional detection is performed, an E_b/N_0 of 12.5 dB is needed to provide the required QoS, which shows about a 6 dB gain over the single antenna case. Furthermore, it is now possible to achieve the desired QoS for conventional receiver at speeds up to 100 km/h. This still does not quite meet the requirement for car receivers (or portable receivers in cars) on highways. Combining the CRC DVBDetect technique with dual antenna diversity brings further improvement. As shown in Fig. 2, the E_b/N_0 requirement is 13.5 dB for receiver speed of 90 km/h, a 4 dB gain over the conventional dual antenna receiver, while at 100 km/h, this gain becomes 8.5 dB. The speed limit is drastically improved to 200 km/h. This guarantees the desired QoS for all car receivers or portable receivers in cars.

It is also worth mentioning that CRC DVBDetect has a relatively low complexity. The additional complexity required to implement this technique into DVB-T receivers is well within the capability of current ASIC and DSP technologies.

5 Conclusions

In this paper, a detection technique is described which can significantly improve the performance of COFDM receivers in fast fading channels. When applied to DVB-T receivers with a single receive antenna, for 8K mode, 64 QAM and rate-1/2 CC, computer simulations show that in a typical urban channels, the speed limit for good QoS (BER of $2 \cdot 10^{4}$ at the Viterbi decoder output) is extended from 90 km/h for conventional DVB-T receivers to 180 km/h for receivers with CRC DVBDetect. For rate-2/3 CC, the speed limit is extended from 30 to 65 km/h. When the proposed CRC DVBDetect technique is implemented in a dual antenna DVB-T receiver, the speed limit for rate-2/3 CC is extended from 100 km/h to 200 km/h as compared to a CCD-based receiver. Therefore, the proposed COFDM detection technique makes DVB-T services with rate-1/2 CC available to all single antenna car receivers (or portable receivers in cars); while it makes DVB-T services with rate-2/3 CC available to car receivers (or portable receivers in cars) with two receive antennas.

6 References

[1] ETSI, 2009. ETSI EN 300 744 v1.6.1, 2009, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television, European Telecommunications Standards Institute, January 2009, 66 pages

[2] ETSI, 2004. ETSI EN 302 304 v1.1.1, 2004, Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H), European Telecommunications Standards Institute, November 2004, 14 pages

[3] FARIA G.: ‘Mobile DVB-T using antenna diversity receivers’, White paper, TeamCast, France

[4] BUROW R., POGRZEBA P., CHRIST P.: ‘Mobile Reception of DVB-T’, White paper, Deutsche Telekom, Berkom, Germany

[5] LEVY Y.: ‘DVB-T – a fresh look at single and diversity receivers for mobile and portable reception’, *EBU Technical Review*, 2004, 10 pages

[6] Commission of the European Communities, EUR 12160 COST 207 Digital land mobile radio communications, 1989, 385 pages

ISSN 2041-5923

Interview – Gustavo Marra

As part of IBC's focus on young professionals working in the industry we present two papers chosen from the poster sessions of IBC 2010. These papers were selected by IBC and the IET as the best young professional poster contributions for IBC 2010.

The first, chosen as the overall best young professional poster contribution is 'Combining digital terrestrial television, GPS system and web digital signage services for vehicles' by G. Marra of TV Globo, Brazil. This paper reports on a system exploiting a location-aware display medium in a public vehicle. This involves using live video when available, stored content when in a reception shadow, and advertising and discount coupons tied to locations on the route of a bus. The paper shows some of the great potential in this area for creative exploitation of technologies through combination.

The second paper, 'Video forensic watermarking on the IPTV STB for traitor tracing' by H.-K. Lee and S.-B. Kim of MarkAny Inc., Republic of Korea, looks at applying watermarking within the existing set-top box environment. Working within the limitations of a piece of hardware can be challenging, but finding ways to utilise limited resources to achieve functionality far beyond the intended limits shows great innovation. This paper takes us inside the 'cat and mouse' world of watermarking/steganography, cryptography and forging practical solutions from sparse resources.

However, before the papers, we present an interview with the author of the best young professional poster contribution for IBC 2010, Gustavo Marra.

How did you get into broadcasting and what do you do currently?

I graduated in Telecommunications Engineering at the University of the State of Rio de Janeiro at the end of 2003. By that time I was already at TV Globo, the most important TV Broadcaster in Brazil and one of the biggest in the world, where I started as an intern in 2002.

Last year I completed the postgraduate course in Networks and Video over IP, at the Federal University of Rio de Janeiro. At the present time I am undertaking an MBA course in project management.

After eight years in the broadcasting market, I am in the position of Project Manager within the Digital Transmission Technologies Department at TV Globo. Our department is responsible for creating and implementing solutions for the contribution and distribution of TV Globo content, and also for the research of new technologies and applications that basically involves transmission and compression of audio, video and data.

What is your paper about?

The paper is about a new type of digital signage service and out-of-home advertising focused on public transportation, where the objective is offering to the passengers (viewers) the most attractive content mixing every possible kind of content delivery available to this environment.

Digital signage, as a method of out-of-home advertising, has been discussed for a long time, but almost all the solutions are based on non real-time content switching between short news and publicity. The main target of most digital signage applications is a short term view, which means that viewers will not spend more than a few minutes in front of the screens, so the communication must be very fast, and the variety and attractiveness of content is not the main concern.

The combination of existing technologies, such as the mobile services of digital terrestrial television, geo-referencing (GPS) and wireless networks (3G), integrated in a video server can bring a different approach for out-of-home advertising using digital signage.

This new approach focuses on high value content for a longer term view, taking advantage of the public transportation service profile, where millions of people spend from minutes to hours every day, available to a more focal advertisement and seeking information and entertainment during their unavoidable journey.

The live broadcasting content offers to the passengers content that could just be accessed at home. The connection through 3G networks can bring the most up to date news and sports results, and GPS can trigger a locally stored advert of a mall that is located in front of the next bus stop.

All this content can be shared together on the same screen, attracting passengers independently who are interested in the match result, the soap opera chapter of the day or the latest breaking news. This attractiveness increases the attention on the digital signage screen, increasing the value of the advertisement.

It is a complete win-win relationship where the public transportation companies can offer a better service to their passengers, TV Broadcasters will keep their viewer even when they are out of home and create new prime times (i.e. rush hours when people are travelling to and from their jobs) and advertisers will have a more valuable tool in terms of attractiveness and the opportunity to focus their message to a specific public or area of the city.

What brought you into this area and what interests you about it?

Digital signage and out-of-home advertising have never been the focus of my work. This relationship came after my deep involvement with the deployment and implementation of digital terrestrial television in Brazil. One important feature of the ISDB-Tb, the Brazilian standard for DTV, is the mobile/portable service called 1-seg. While studying possible applications for the new mobile service, the idea emerged. After some discussions with a traditional digital signage services provider from Brazil, the drawing of the system was completed.

A characteristic of this work that is very attractive for me is the possibility of combining existing technologies to create a new service, capable of reaching millions of people every day, increasing the value of content and advertisements.

In your opinion, how will digital signage and out-of-home advertising develop from here?

As new technologies come through, new possibilities will be available for digital signage and out-of-home advertising. The system suggested in the paper brings the possibility of a focused kind of advertising, where the characteristics of a specific bus or train route can be explored through publicity. The focus could be specific stores or brands that are present or interested in a specific neighbourhood.

What I imagine for the future of this area is the constant search for a more directional publicity, with more specific and targeted ads, associated to valuable content. If today we can already talk directly to specific groups with habits and behaviours in common, in the future we will be able to talk to specific people, filling their expectations exactly.

What do you think may prove to be the biggest challenge in achieving this vision?

I see two important challenges to achieve this vision of the future: technology and understanding. To be able to offer a focal advertising and high value content on digital signage screens, solution providers must have a great understanding of what viewers want on each kind of application and offer the best content for that environment. Regarding the specificity of the advertisement, it is still missing a technology that allows the system to get in touch with every specific person, advertising products related to their specific profile and lifestyle, and showing the most interesting content for each one.

Is this the first paper you have submitted to IBC and have you been to the conference before?

Actually it is not only the first paper I have submitted to IBC, but my first paper ever submitted. Almost all my research is directed to solutions or applications related to my area of work and the result is a project or a new feature or implementation to an existing system. After completing the studies on this system I realised that this research could also result in a technical paper, mainly because of its originality in creating a new service combining existing technologies.

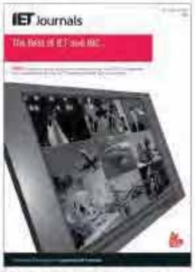
Last year I visited the IBC exhibition for the first time. The visit didn't have a specific focus; my objective was finding new technologies and solutions on signal compression and transmission. The result was very positive and I was able to get in touch with technology developers in very high level discussions. I also had good experiences with the open session conferences.

Apart from presenting your poster, what else will you be doing at the conference? Are there any sessions you are particularly interested in?

After reading the IBC 2010 Conference Programme I realised that it will be difficult to organise the time to reach the biggest number of sessions as possible. The sessions' subjects are really attractive.

As I am concentrating my research mainly on 3D and broadband TV applications this year, those will be the topics that I will focus on at the conference.

Selected content from IBC 2010. This paper was voted best young professional poster contribution



ISSN 2041-5923

Combining digital terrestrial television, GPS system and web on digital signage services for vehicles

G. Marra

Digital Transmission Projects, TV Globo, Rua Lopes Quintas 303, Sala 301 – Jardin Botanico, Rio de Janeiro, Brazil
E-mail: gustavo.marra@tvglobo.com.br

Abstract: A new type of digital signage solution for different kinds of public transportation, such as buses and trains, is suggested. The system takes advantage of the portable service (1-seg) available on the digital terrestrial television signal in Brazil and internet connection through mobile phone networks to provide the most attractive and updated content of news and entertainment. Combining a local video server and a global positioning system (GPS), the system can provide targeted advertisement to specific public on each route of public transportation and also avoid terrestrial coverage shadow areas.

1 Introduction

A new way to provide digital signage has just been developed within TV Globo. This new system combines indoor advertisement and entertainment on vehicles, such as buses, trains, subways and even airplanes, with live broadcasting in a very compelling way. It combines live broadcasting, web (3G networks), geo-referencing and video-server technologies. The services offered in the vehicles through this new system combine these different technologies to show the passengers (viewers) the most attractive content possible.

On the digital signage screens inside the vehicles, different kinds of contents are shown, combining live broadcasting content (ISDB-T 1-seg) [1], content from the web such as news and sports received by 3G networks, locally stored advertisement, static advertisement and programming teasers (shown as banners). The arrangement of contents on the screen can be totally customised by the user.

The geo-referencing (global positioning system, GPS) is used for two purposes: (1) to couple with the few existing shadow areas of the broadcasting signal; (2) to trigger location based advertisements. As these vehicles have a known route, these shadow areas can be avoided by the system replacing live broadcasting for local content, triggered by GPS. The geo-referencing is also used to trigger advertisements based on the position of the vehicle: the system may play advertisements of a store that the vehicle is just passing by.

The system allows automatic replacement of the live broadcasting commercial adds by a locally stored advertising clip, with the same duration of the live break. This feature creates new prime times for advertisement in rush hours.

2 System/services description

The system suggested in this paper has the following composition:

- Robust and small sized CPU – video server
- ISDB-Tb 1-seg vehicular receiver and antenna
- 3G modem and antenna
- GPS receiver and antenna
- FM transmitter for audio
- Vehicular displays

These components are illustrated in Fig. 1.

The system can be described as a video server/storage that combines different inputs of the system and formats the video displayed on the screen. Live content is provided by the ISDB-Tb 1-seg receiver, which delivers an audio and video signal to the server while web access is provided by the 3G modem, using the mobile phone operators' network. Geo-referencing is provided by the GPS receiver and data is processed by software at the server. An FM transmitter is used for audio transmission in the vehicles and a Bluetooth transmitter is used to distribute audio and video content, coupons or music to mobile phones.

The system delivers live broadcasting content, web content with real-time actualisation and locally stored content to the digital signage display. To improve the specificity of the service, locally stored content and ads can be triggered by GPS or time (i.e. a mall or a specific store can be advertised a few miles before the public transportation passes nearby). Web content can be used for breaking news, sports news and entertainment content and can also be triggered by geo-referencing using GPS or time (i.e. the traffic condition on the bus route can be updated combining web access and GPS). Using Bluetooth, movie teasers, advertisement contents and coupons can be sent to mobile phones. Live broadcasting audio is transmitted using an FM modulator, so viewers can listen to it using mobile phones or any FM receiver.

The layout of the display where the content is shown can be rearranged anyhow, depending on the kind of application or the creativity of the service provider. As it is the most premium content available on the platform, live broadcasting content is suggested to occupy the biggest part of the screen division, as shown on Fig. 2.

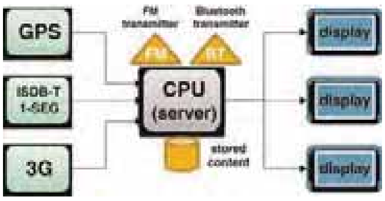


Figure 1 *System diagram*

3 Scenario description and market size

The service was developed for the use on public transportation, such as buses and subways, in cities where digital terrestrial television is available in Brazil. There are currently 39 cities with digital terrestrial television coverage in Brazil [2], including major cities such as Rio de Janeiro, Sao Paulo and Brasilia, and it is expected that 45 cities will be covered by the end of 2010 [3].

There are almost 15 000 vehicles in use for public transportation only in the city of Sao Paulo [4]. In the year of 2009, 2.9 billion passengers used public transportation in the city of Sao Paulo [5]. If the system is installed in 30% of these public transportation vehicles it means that almost one billion passengers would become viewers in one year, in a single city.

The target clients of the system are the public transportation operators such as bus companies, subways and airlines. This service will enable the increment of advertising revenues to these companies, considering that advertising is already provided by transportation companies as static banners and conventional digital signage.

The target viewers of the service are the users of public transportation services in Brazil in cities covered by the digital terrestrial television service.

4 Advantages and opportunities

If compared to usual digital signage services and static advertisement (banners), the main advantages of the presented system are live content (broadcast and web), which will increase the attractiveness of the service.

In Brazil, most of the premium content is available free-to-air, like blockbuster movies and major sport events, which makes the 1-seg live content very attractive.

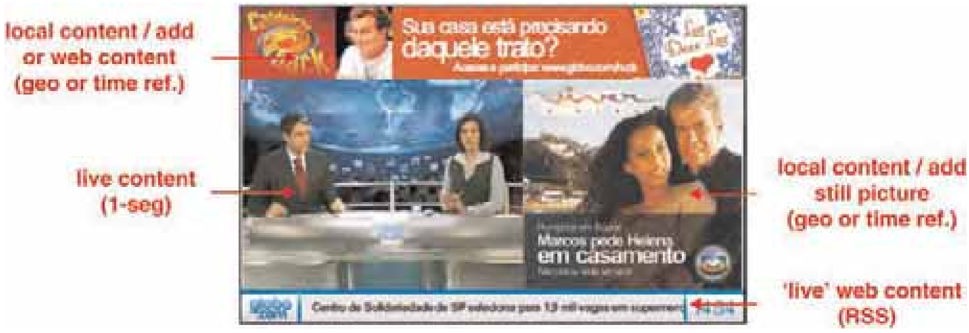

local content / add or web content (geo or time ref.)
live content (1-seg)
local content / add still picture (geo or time ref.)
'live' web content (RSS)
Sua casa está precisando daquele trato?
Marcos pede Helena em casamento

Figure 2 *Service display: layout option*

Regarding the advantages on out-of-home advertising, geo-referencing can be used for focal and specific advertisement, i.e. malls and stores may be advertised close

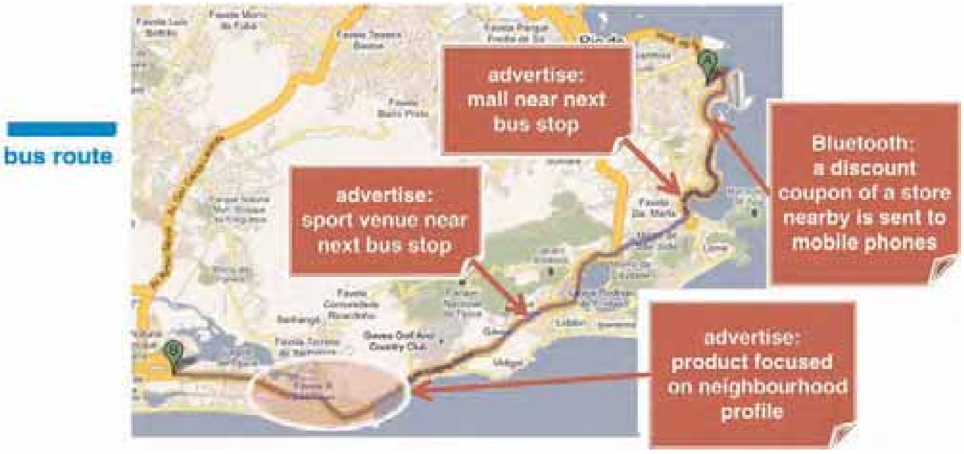

bus route
advertise: mall near next bus stop
Bluetooth: a discount coupon of a store nearby is sent to mobile phones
advertise: sport venue near next bus stop
advertise: product focused on neighbourhood profile

Figure 3 *Advantages and opportunities provided by the solution*

to their location, a product may be advertised in a specific neighbourhood crossed by the vehicle route or Bluetooth may be used to transmit teasers, ads and coupons to mobile phones, which also can be triggered by GPS.

These advantages and opportunities are illustrated in Fig. 3.

5 Main challenges and issues

The main challenges and issues of the service are certainly related to legal issues, digital terrestrial television coverage and display video quality and robustness.

Regarding the legal questions, an important issue is that different cities have different legislation regarding the video services in public transportation. A very deep study is required to meet all the legal barriers using the same system.

Talking about signal coverage, depending on the geography of the city, the digital terrestrial television coverage can present shadow areas where the 1-seg signal is not received by the vehicle. As the public transportation vehicles have pre-known routes, live broadcasting video can be replaced by locally stored content using GPS.

Finally, with regards to display video quality and robustness, some manufacturers already provide vehicular monitors with high brightness and contrast, and enhanced robustness that can be considered 'vandalism proof' [6].

6 Conclusions

The system/service is ready and is expected to be launched this year in Rio de Janeiro, Brazil. The commercial use involves understandings between public transportation companies, system/hardware provider and broadcasters.

The system can be used in countries that provide different mobile broadcasting standards, replacing the ISDB-Tb receiver for DVB-H, Flo, ATSC-MH, DMB and other standard receivers. It means that the digital terrestrial standard that supports mobile/portable services could be used for live broadcasting.

The service presents a win-win relationship, increasing advertising revenues (for transportation companies), a better return on investment for target ads (for announcers), an increment of out-of-home viewers and new prime time hours (for broadcasters), and much more appealing entertainment and information content for passengers.

7 References

[1] Brazilian Forum of Digital Terrestrial Television, Standards – http://www.forumsbtvd.org.br/materias.asp?id 112

[2] Brazilian Forum of Digital Terrestrial Television, Implementation schedule – http://www.forumsbtvd.org.br/materias.asp?id 55

[3] Brazilian Forum of Digital Terrestrial Television, Clipping – http://www.forumsbtvd.org.br/materias_index.asp?menu 5

[4] Sao Paulo transportes SA – SPTrans, Indicators of the fleet – http://www.sptrans.com.br/indicadores/ – click on *frota* (fleet)

[5] Sao Paulo transportes SA – SPTrans, Indicators of transported passengers – http://www.sptrans.com.br/indicadores/ – click on *passageiros transportados* (transported passengers)

[6] Technovision, Vehicular LCD displays – http://www.technovision.com.br/monitores-lcd-veiculares-onibus

Selected content from IBC 2010. This paper was voted an outstanding best young professional poster contribution

ISSN 2041-5923

Video forensic watermarking on IPTV STB for traitor tracing

H.-K. Lee S.-B. Kim

MarkAny Inc., IOF Ssangnim Bldg. 151-11, Ssangnim-dong, Jung-gu, Seoul 100-400, Republic of Korea
E-mail: hklee@markany.com

Abstract: A real-time invisible forensic watermarking method on an internet protocol television (IPTV) set-top box (STB) without considering a human visual system model is proposed. This method uses an overlay (graphics) plane for watermark embedding because most STBs do not provide an external interface to access a video frame. For real-time embedding, a pre-processing procedure and an efficient watermark embedding structure are employed. The watermark embedding into a graphic plane has an effect on the variation of the luminance intensity because the watermarked graphic plane is finally blended with a video plane according to the alpha source compositing rule to display a video on a TV screen. The robustness of the proposed method is proven by a variety of experimental evaluations and analysis on real STB environments.

1 Introduction

The digital video contents service on internet protocol televisions (IPTVs) has made our life more convenient and controllable because we can see a specific movie or drama whenever and wherever we want. Meanwhile, a new challenging issue has occurred: that is illegal capturing and distribution of a copyrighted video. A user could easily record a transmitted video image from a set-top box (STB) using a capture device or a camcoder for displaying it on an IPTV screen. Herein, a capture device equipped on a personal computer connected to an output channel, such as s-video, or component channel of a STB. The capturing is done after the cryptographic decoding of a video protected by a conditional access system digital rights management (CAS-DRM) is completed. Thus, currently, there is no perfect protection scheme that exists to prevent a user from capturing and camcording.

In addition, the cost of the capture device has recently decreased while its performance has been powerful. It includes a hardware based encoding chip that enables a user to effectively save the video frame into the specific video codec without frame delay. A portable multimedia player (PMP) device also has a capturing function and a composite channel. Thus, a user could easily capture a video that is transmitted from a STB to a TV. From this fact, a countermeasure for copyright ownership proof such as the digital watermarking or feature based video identification technique is required. The feature-based video identification technique, such as video fingerprinting and video copy detection [1, 2], has received increasing attention for contents filtering. This technique requires large amounts of the fingerprinting database to search the queried contents. In addition, we could not obtain illegal distributor information using this technique.

Owing to the above challenging issue, the watermarking technique is considered as an alternative way to prevent a user from uploading the captured video through the internet and obtain illegal distributor information. Herein, there are some issues to be considered. A STB not only has low computational resources but also does not allow access to a decoded video frame. From these restrictions, we consider a simple and fast watermark embedding algorithm for real-time forensic marking. In addition, we consider a video watermark embedding algorithm that does not use a human visual system (HVS) model. A HVS model is used to invisibly embed a watermark into a video frame. To resolve these issues, the proposed method uses a graphics (overlay) plane that is a watermark target plane. Herein, the modulated watermark image is just drawn on a graphics plane that will be blended with a video plane according to the alpha-composite rule [3].

This report will begin with the problem definition and a description of the technology. Then, a fast and simple watermark embedding algorithm will be described, and the experimental evaluations will be shown. Finally, discussions about the advantages and disadvantages of the proposed method and concluding remarks are described.

2 Background

The digital video contents service has recently spread into our life owing to an interactive IPTV service that provides the video on-demand (VOD) service as well as real-time broadcasting service. Therefore, a user could see a movie or a drama at any time that they want. They also could see recent movies directly with easy payment. By doing this, the user can select a movie that he/she wants to see. Then, the selected movie is generally downloaded into a STB which is a front-end device for playing the downloaded video. Concurrently, the CAS-DRM protection mechanism decodes the encrypted video, and then the decrypted video frame is transmitted into a digital displaying device such as the TV via component channel, for example. At this time, s-video, high definition multimedia interface (HDMI), composite, and component channels can be used as the connection to an input channel of a displaying device. The HDMI interface has a lock, such as the copy protection mechanism [4, 5], while s-video, composite, and component channels are a bit free because the capture device only checks these channels once in first playing time. Thus, a user who has a capture device could record and save a video via these channels into the digitalised one. Then, illegally captured video can be distributed through the Internet. It might give serious commercial losses to content owners and providers.

The traditional video watermarking technique uses a HVS mask to minimise a visual distortion occurred from a watermark embedding. In addition, the watermark can be embedded on a spatial domain or transform domain as well as on a compressed domain or an uncompressed domain. The choice of the embedding environment can be selected according to the required applications. For real-time service, the compressed domain watermarking method can be considered, but there is a problem that the robustness against various attacks is reduced. In addition, the calculation of a HVS mask needs high computational costs and is time consuming task. Thus, a development of a simple HVS mask is required.

For these reasons, the traditional video watermarking approach is not pertinent to an IPTV STB environment. Most STBs do not provide a direct access to a video frame as well as a video plane. These restrictions do not allow the use of a HVS mask as well as direct modification for watermark embedding. Thus, the new watermarking scheme is required. To resolve the above issues, a graphics plane for a target watermark region is selected, and then a watermarked graphics plane and a video plane are finally blended which makes a watermark reflects to a final TV screen. The effect of the watermarked screen is the brightness change. To reduce a computational cost and provide fast watermark embedding, a pre-processing task is employed. Fig. 1 describes the proposed overall watermark embedding procedure on an IPTV STB.

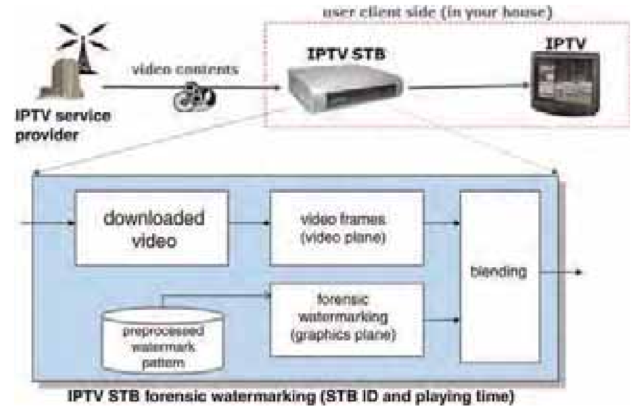

user client side (in your house)
IPTV service provider
video contents
IPTV STB
IPTV
downloaded video
video frames (video plane)
blending
preprocessed watermark pattern
forensic watermarking (graphics plane)
IPTV STB forensic watermarking (STB ID and playing time)

Figure 1 *Forensic watermark embedding procedure on an IPTV STB environment*

3 Proposed video forensic watermarking on IPTV STB

The real-time requirement for watermark embedding is an essential factor on an IPTV STB to embed a user tracing information. In addition, general STBs have low memory and computational resources. Thus, simple and efficient watermark embedding methods are required. To cope with above issues, a pre-processing method off-line and pattern drawing with the pre-processed watermark pattern are employed for forensic watermarking as shown in Fig. 1. Detailed methods are as follows.

3.1 Forensic watermark embedding

We embed 83 bits payload representing a STB ID and the playing time as a forensic mark in real-time on a STB. An illegally captured video from a STB is mostly distorted by DA/AD conversion, scaling, frame rate change, compression, and translation. Thus, finding correct geometrical parameters such as rotation, scale, and translation (RST) is an important task to reliably extract the embedded watermark message. Since we embed a generated watermark into a graphics plane without a HVS mask as well as with a minimum embedding strength in order not to reduce the visual quality, a periodic watermark pattern for detecting geometric parameters (RST pattern) and a message pattern are separately embedded in timeline as shown in Fig. 2.

Each watermark pattern uses spread spectrum based embedding method on a spatial domain. The RST pattern is with 4 × 4 grid blocks as shown in Fig. 3. The resolution of a graphics plane is 960 × 540. The size of the RST pattern is 80 × 55 which is up-sampled two times.

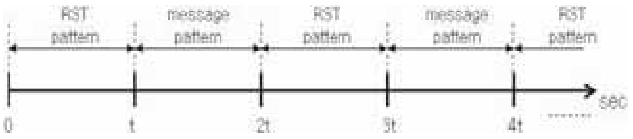


Figure 2 *Watermark embedding sequence on a graphics plane*

Herein, the m-sequence [6] with 1024 length and a secret key is employed for watermark signal generation. The small margin at the border area is remained unmarked because this region may not be displayed on a real TV screen dependent to a device driver software when playing a video. The 4 × 4 grid blocks with 160 × 110 sizes are used for watermark embedding. Meanwhile, the message pattern is with 2 × 2 grid blocks as shown in Fig. 4. The size of the message pattern is 160 × 110 which is up-sampled two times. Herein, the m-sequence with 1024 length and a secret key is employed for watermark signal generation. The m-sequence with 1024 length can represent 10 bits. Thus, 170 bits (1024 × 17   17 408) can be embedded into a 160 × 110 block (17 600 pixels). 83 bits payload, additional information 27 bits, and error correction code 60 bits are modulated. The Reed–Solomon code [17 11] for error correction [7] is employed. Its false alarm rate is 6.32E-7.

The resolution of a graphics plane is fixed, and corresponding RST pattern size is also fixed as well as not changed. Thus, the fixed 496 × 904 RST pattern image is generated off-line, and then it is just drawn on a graphics plane according to the time interval *t* as shown in Fig. 2. The message pattern is generated before an on-demand video is played since a commercial advertisement is mostly broadcasted while an on-demand video is buffered for playing. Thus, the message pattern has also only to be drawn on a graphics plane according to the time interval *t* as shown in Fig. 2. The RST and message patterns are repeatedly drawn in turn. At this time, the alpha-compositing rule [3] is applied on a STB. This rule is to

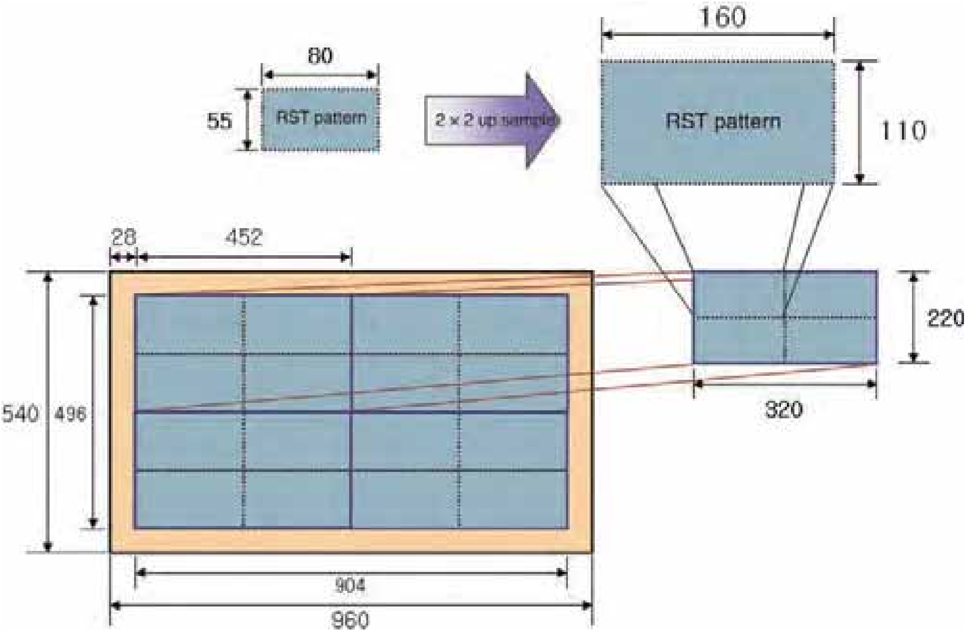


Figure 3 *RST message structure*

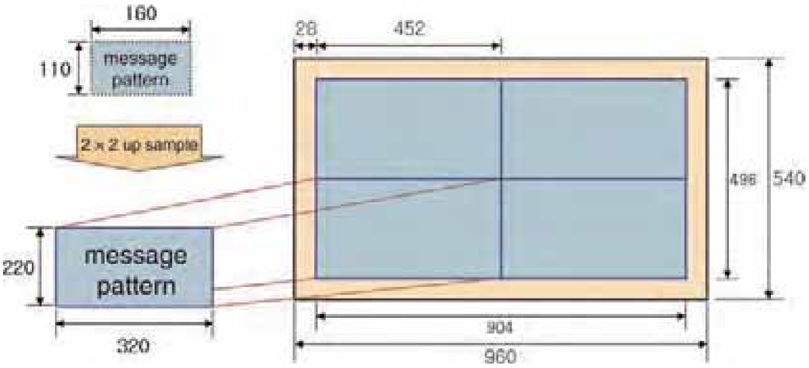


Figure 4 *Watermark message structure*

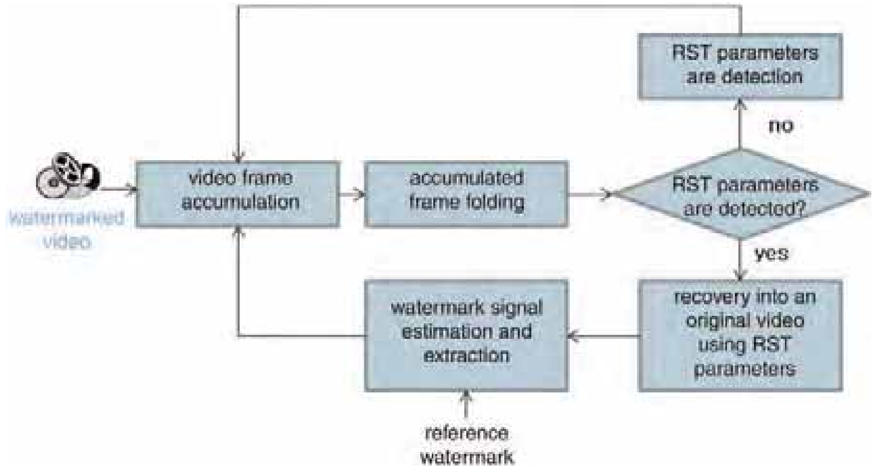

watermarked video
video frame accumulation
accumulated frame folding
RST parameters are detected?
no
RST parameters are detection
yes
recovery into an original video using RST parameters
watermark signal estimation and extraction
reference watermark

Figure 5 *Watermark extraction flow*

blend some images on a graphics plane, and then blend a composite image of a graphics plane with a video frame of a video plane. From the previously generated image, we have only to draw a watermarked image according to a decided time interval. This method is very fast and needs low computational costs. Thus, this method enables the real-time watermark embedding on a STB.

3.2 Watermark extraction

To extract an embedded watermark, a blind watermark detector is performed by normalised cross correlation. The first step is to find the geometric parameters. To do this, autocorrelation function of an estimated watermark signal is performed, and then periodic peak signal is estimated which is owing to the fact that we embedded a periodic signal. Adaptive Wiener filter with a 3 × 3 window as a de-nosing filter is employed to estimate a watermark signal. The geometric parameters, such as rotation, scale, and translation, are obtained using a method [8] from a detected peak image.

The next step is to extract an embedded watermark signal with the detected geometric parameters. The watermark signal is accumulated and folded into a 220 × 320 image to enhance the estimated watermark energy during $t/2$ seconds. Herein, the video frame is recovered into an original resolution by using the detected geometric parameters. Then, normalised cross correlations between the reference watermark sequences with a secret key and an estimated watermark signal is computed where the index having a maximum correlation value is selected. The constructed bits with the selected indexes are verified by a Reed–Solomon code. Fig. 5 describes a watermark extraction flow.

4 Performance evaluations

4.1 Test environments

The proposed method is implemented and evaluated on the real STB environment where the chip-set for decoding a video is [9]. The performance of the proposed method is evaluated using 10 movies wherein these movies have a variety of sub-sequences including sports, landscapes, action, and animations. The total length of the movies is approximately 10 h. The base modification as a test type-I are DA/AD conversion into 2 Mbits, MPEG-2 video, resized from high definition (HD) to standard definition (SD) resolution, and frame rate change from 29 ~ 25 to 30 fps. The test type-II is a format conversion from MPEG-2 to MPEG-2, MPEG-4, H.264.AVC, and re-compressed from 1 to 2 Mbits. The test type-III is a slightly rotated set by a multiple of 1° in between 5° and 5°. The test type-IV is a letter-box and a pillar-box inserted set. The test type-II, III, and VI are all re-sampled once again using a test type-I set. Thus, total 40(10 × 4) query videos are used to show the extraction performance.

4.2 Robustness

To show the robustness against a variety of video distortions, videos are transmitted into a test STB from a media transmitting server. The proposed forensic watermarking module embeds the user tracing information into a graphics plane where the received video on a video plane is blended with the watermarked graphics plane and then displaying it on an IPTV screen. Actually, the blended final video signal is transmitted from a composite channel of a STB to an input channel of an IPTV.

To generate test videos, we captured and recorded the received video on a STB where a composite channel on a STB is connected to a capture card on a personal computer. As a capture card, The OnAir capture board [10] is selected because it is inexpensive and includes the hardware based MPEG-2 chip. The captured videos are saved into a 30 fps, 6 Mbits, SD (720 × 480) resolution, MPEG-2 video type. The length of each video is about 16 min where 12 numbers of forensic watermarks are embedded. The captured video has modifications that are DA/AD conversion, frame rate change, codec conversion, and scale-down. We extracted an embedded forensic watermark from the captured videos. We successfully extracted all of embedded forensic watermarks as shown in Fig. 6. The captured video is generally recorded with low compression to fully capture a video frame. Thus, a user generally re-encodes a captured video to have a small sized and high quality video. From this reason, some experimental evaluations against transcoding attacks are performed.

First, the compression attack experiment is evaluated. The captured 6 Mbits, MPEG-2 video is re-compressed into from 2 to 1 Mbits MPEG-2, MPEG-4, and H.264/AVC videos respectively. The resolution is changed from SD to SD and from SD to CIF as shown in Fig. 7. Fig. 7 showed the extraction accuracy when a forensic watermark is embedded 12 times into a video. The performance of the proposed method showed that the watermark extraction is robust against even to 1 Mbits compression attack in both SD and CIF resolution. In one video, it described that the extraction success rate more than 50% of embedded watermarks is possible.

Next, the frame rate change, letter-box and pillar-box insertion attacks are evaluated as shown in Fig. 8. The regenerated videos are 2 Mbits, MPEG-2 videos. The frame rate change is one of the frequently occurred attacks where we changed a frame rate from 30 to 25 fps. Fig. 8 showed that the proposed method showed the good performance in a frame rate change than the one of compression attack in both SD and CIF videos. The letter-box and pillar-box can be included into a captured video because these are included when an aspect ratio is changed. Herein, we compulsorily insert 25% of letter-box and pillar-box into a video, respectively. Fig. 8 showed that the proposed method is robust against letter-box and pillar-box insertion attacks.

The rotation attack is not a frequently occurrence in a video. However, this attack can be performed with an intentional purpose by a user. We re-generated the

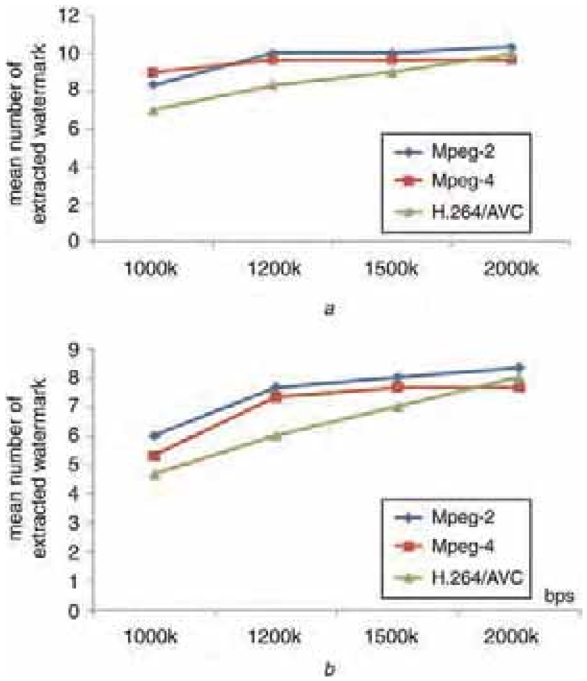


Figure 7 *Robustness against compression attacks: test set type-II*

a Robustness against compression after capturing a video into SD resolution

b Robustness against compression after capturing a video into CIF resolution

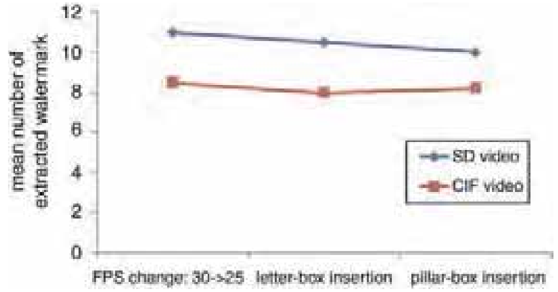


Figure 8 *Robustness against frame rate change and aspect ratio change attacks: test set type-IV*

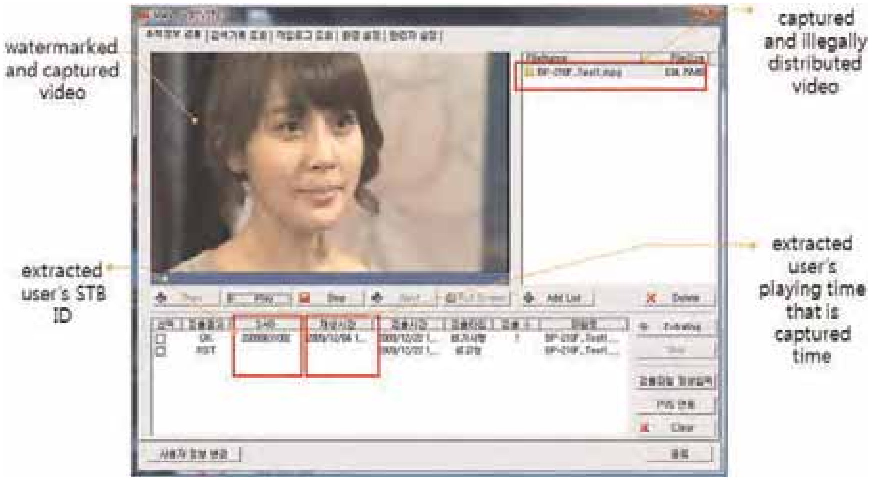

watermarked and captured video
captured and illegally distributed video
extracted user's STB ID
extracted user's playing time that is captured time

Figure 6 *User interface for watermark extraction: watermark extraction example*

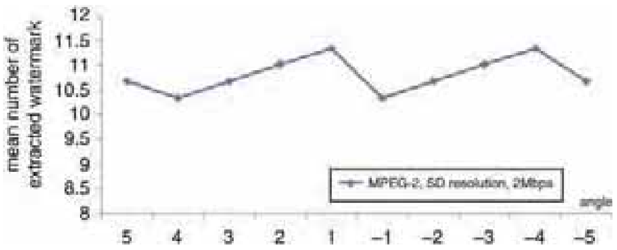


Figure 9 *Robustness against rotation attacks: test set type-III*

captured videos into a 2 Mbits, SD resolution, MPEG-2 videos that are rotated in between 5 and 5 angles as shown in Fig. 9. Fig. 9 showed that the proposed method is robust against slightly rotated video distortions.

The embedding strength of our proposed method is fixed with a uniform constant value because we can not compute a HVS mask. In addition, the embedding strength is minimum value to reduce a visual quality. From these reasons, the watermark extraction accuracy can not be high when we just use a few video frames. To increase the watermark extraction performance, the video frame accumulation and folding method is used. The more the number of video frames for accumulation and folding is used, the more performance is increased because the watermark signal is strengthened. Experimentally, 40 s of video frames are pertinent to the attack environments, such as DA/AD conversion, compression, scale-down attacks, with respect to the robustness and the size of the required minimum watermark segment. In our proposed method, we embedded the RST detection information and real watermark payload, respectively. Thus, total 80 s of video frames are required to extract the one of the watermarked information. To increase the robustness against other attacks such as camcorder recording, the more video frames should be embedded and then accumulated for watermark extraction.

5 Conclusions

The proposed method provided a simple and fast watermark embedding method that is practical in the commercial STB environment. Since the HVS mask is basically not used to maintain the visual quality of a video, a watermark embedding strength should be minimised where the watermark signal strength is weakened. To cope with this issue, all watermark payload bits are embedded into one video frame, and then these are repeatedly embedded into a consecutive video frame during a given *t* s since the accumulation and folding of consecutive video frames make a watermark signal strong in the watermark extraction procedure. The proposed method is evaluated with the captured videos from a STB and their geometrically distorted videos after DA/AD conversion. The evaluation results showed the good robustness on our test environments. However, we think that the extraction algorithm must be compensated to cope with a captured video by a camcorder. We think that this is not easy work because very weak watermark signal is embedded into a video. We think that the complete solution is that the watermark embedding module should be incorporated into a decoder chip of a STB. However, to cope with the already installed STBs house, we think that the intermediate solutions, such as our proposed method in a user's, will be used before complete solution is developed.

6 Acknowledgments

This research was supported by the Korea creative content agency (KOCCA) where its project number is 2-09-1205-001-10987-09-001.

7 References

[1] SUNIL LEE, YOO C.D.: 'Robust Video Fingerprinting for Content-Based Video Identification', *IEEE Trans. Circuits Syst. Video Technol.*, 2008, **18**, (7), pp. 983–988

[2] ARUN HAMPAPUR, KI HO HYUN, RUUD BOLLE: 'Comparison of Sequence Matching Techniques for Video Copy Detection'. Proc. SPIE, San Jose, CA, USA, 2001, vol. 4676, pp. 194–201

[3] THOMAS PORTER, TOM DUFF: 'Compositing digital images'. ACM SIGGRAPH Computer Graphics, 1984, Vol. 18, no. 3, pp. 253–259

[4] DOËRR G., DUGELAY J. L.: 'A guide tour of video watermarking', *Signal Process., Image Commun.*, 2003, **18**, (4), pp. 263–282

[5] LINNARTZ J. P.: 'The ticket concept for copy control based on embedded signalling'. Proc. Fifth European Symp. on Research in Computer Security, Louvain-la-Neuve, Belgium, 1998, (*LNCS*, **1485**), pp. 257–274

[6] SASAKI Y.: 'M-sequence generator and PN code generator with mask table for obtaining arbitrary phase shift'. United States Patent 6339781, January 2002

[7] LIN S., COSTELLO D.J. JR.: 'Error control coding: fundamentals and applications' (Prentice Hall, Englewood Cliffs, NJ, 2004, 2nd edn.)

[8] KARYBALI I.G., BERBERIDIS K.: 'Efficient spatial image watermarking via new perceptual masking and blind detection scheme', *IEEE Trans. Inf. Forensics and Security*, 2006, **1**, (2), pp. 256–274

[9] STi7100: Low cost HDTV set-top box decoder for H.264/AVC and MPEG-2, http://www.st.com/stonline/products/literature/bd/11102/sti7100.pdf

[10] OnAir website, available: http://www.onairsolution.co.kr/

ISSN 2041-5923

Introduction to *Electronics Letters*

Electronics Letters[1] offers researchers a unique avenue to rapidly disseminate significant work to a large international audience in a short paper format, enabling its readers to be kept abreast of the very latest developments. *Electronics Letters* encompasses a very wide range of interdisciplinary areas covering nearly all aspects of electrical and electronic technology from work relating to the materials used to create circuits, antennas and other components all the way through to software. Some areas are obviously and directly relevant to broadcast media including image and video processing, communications technology and display hardware; other areas also have indirect relevance as their research will form the foundations of the broadcast media technologies of the future. In the Letters selected to appear here we have concentrated on the former – work which has immediate and direct relevance to IBC 2010 attendees, addressing current and near future issues and technology in broadcast media.

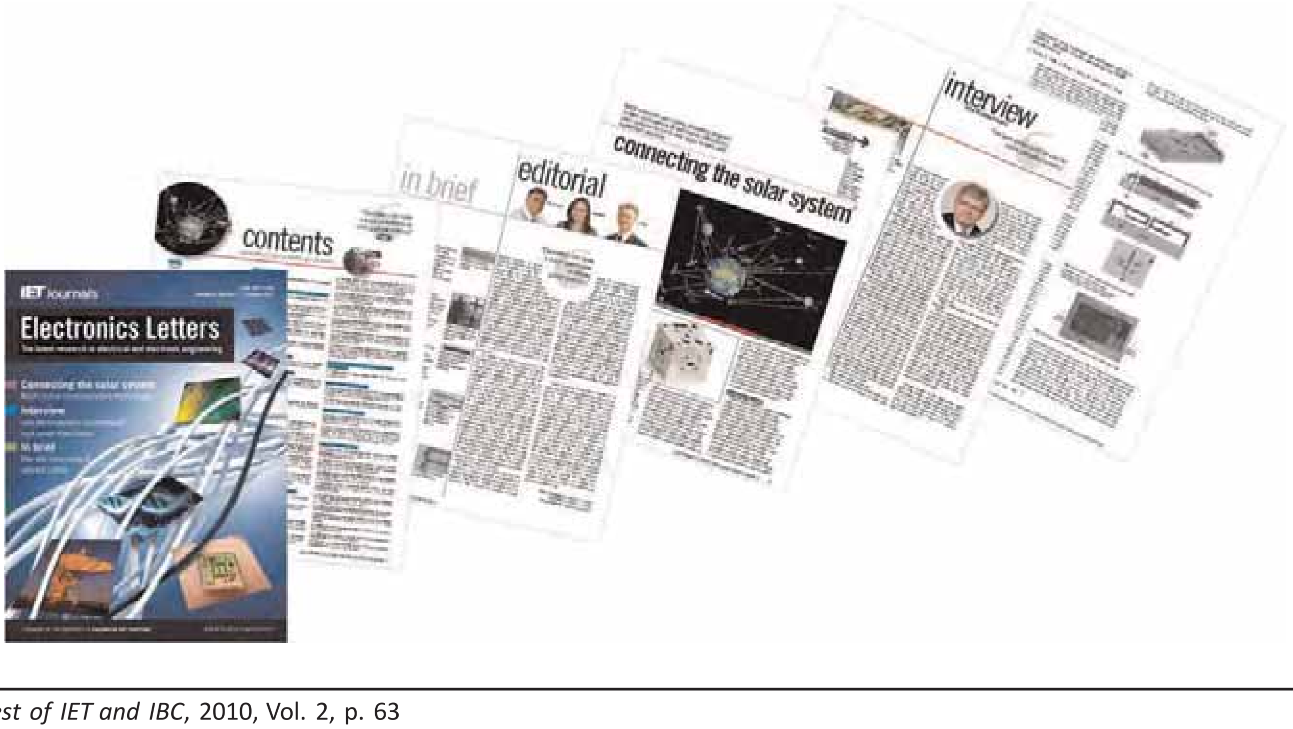
First, we have an interview with the European Broadcasting Union's David Wood, Deputy Director of EBU Technical, talking about 3D-HDTV, how he got into the field and what he sees in its future. This interview, part of the new format *Electronics Letters* which now includes colour articles providing context for selected papers, was conducted to complement an invited paper David had written for *Electronics Letters* exploring the current state and potential evolution of the field as part of our Insight Letter[2] series. In this series, experts from different fields share their thoughts and opinions about the current state and possible futures of their research areas. David Wood's Insight Letter 'Model behaviour for 3D-HDTV' is also included and follows the interview.

We hope that these papers will provide you with an insight into the kind of broadcasting-relevant, new and interesting research published in *Electronics Letters*.

The *Electronics Letters* editorial team.

[1] www.theiet.org/eletters

[2] All our Insight Letters are free to view and can be found at www.theiet.org/insightletters



Interview with David Wood

How did you come to work in 3DTV research?

I'm an 'old timer'. I was researching 3DTV in the 1980s. We trialled a system in Germany, an 'anaglyph' system with red and cyan as the complementary colour casts of the L and R pictures, combined in the broadcast PAL picture. We sold (yes, sold) over 6 million pairs of glasses for the first live broadcast from the NDR studios in Hamburg. The viewers saw a monochrome 3D picture, full of maids poking feather dusters into the viewer's face – very cultured. Being monochrome in a time when colour TV was blooming, it really wasn't going anywhere, so we didn't do more live broadcasting. But we did make local trials with a colour system based on two PAL projectors and polarised glasses, to get the reactions of viewers[1].

After that, other things went to the top of the European Broadcasting Unit's (EBU) technical agenda, principally high definition TV (HDTV), so my own work turned to that. But I never forgot 3D, and became a private 3D still photographer. I really was chuffed when 3DTV returned to broadcast studies several years ago.

In your Insight Letter you envision three generations of 3D technologies with current stereoscopic systems in the first generation. What do you think will be the main challenges for the later generations?

The main second generation barrier is probably the horizontal resolution of the display. It will be 'auto-stereoscopic' (no glasses). Displays need some means, such as a lenticular surface laid over the screen, to ensure each eye sees a different picture, via 'columns' that make up the picture. But each column included eats up the screen's horizontal resolution, so that's the dilemma. The more columns (and views) you have the better in terms of more parallax, but the worse in terms of horizontal resolution. Getting that right with a large screen display is the challenge for the second generation. It's similar for the third generation except the dilemma occurs vertically as well as horizontally.

The 'object wave' recording of the later generation will be a fantastic challenge. A hologram records an object wave, but it's a simple one. We are at a kind of 'AM' equivalent stage in light wave modulation. The spirit that moved engineers from AM to FM to digital modulation and compression in radio waves needs to be applied to optics, and surely will be.

Speaking of which, in your evolution model the third generation allows 'natural vision', what does this mean and what will it require?

Natural vision comes when you are able to capture, record and transmit an object wave. Imagine holding up an empty picture frame in front of your face. What passes through the frame on its way to your eye is the object wave. Somehow we need to be able to record the amplitude, phase, and frequency of that wave. We probably need a way to modulate a reference wave with the object wave which folds in the phase information to the amplitude information, with three sensors for red, green, and blue, and then compression will be needed. The amount of information is going to be huge before compression, if we tried to broadcast it today, one channel would take up the entire broadcast bands.

What excites you about this field?

The fantastic thing about 3D, done well, is that it's like a time machine. You capture a moment in time forever. It gives you goose bumps when you see, say, a large transparency 3D format in ideal viewing conditions. To know that eventually we will achieve this with television – fantastic. It's also fascinating to work in an area that non-scientific people think is close to magic.

Given its history, do you think 3D's time has finally arrived – what is different this time around?

Realistically, the cinema world likes 3D because it brings people back into the cinema, with expensive popcorn, and away from their spanking new HDTV at home. Nor can we imagine that the set makers are not looking for new things to sell us, with sales of flat panels flattening off. But, in addition, there is science. Today we have digital processing that can be applied to the left and right pictures, to make sure the registration is perfect – and we have large HDTV flat panels instead of standard definition TV (SDTV) and cathode ray tubes (CRTs). There's never been a time when 3DTV had as much of a chance as it does today.

Apart from the electronics, what needs to be done to make 3DTV successful?

There are two things. First, we have not done nearly enough psycho-physical stuff, evaluating eye-discomfort and eye-fatigue. We don't have enough scientifically derived results, only anecdotal or small scale evidence. We need to know, particularly, if and how it could affect young people whose eyes are still learning and growing. It's the academic world that needs to do this, on behalf of us all. I hope they are listening!

Secondly, it's not enough for engineers to develop the 3D broadcasting technology. For it to be successful, there must be programmes which make good use of it. Creative thinking is needed about what could 'suit' 3D in a TV programme. I remember in the 1980s discussing with programme makers what we could do with 3DTV. There was a popular quiz show at the time called 'Celebrity Squares'. One producer said 'What about a "Celebrity Cubes"?'. Well, we need better ideas than that this time.

[1] Some of David Wood's early 3DTV work is found in Messerschmid U., Sand R., Wood D.: 'Relief in sight?', *EBU Tech. Rev.*, 1986, **37**, (6), pp. 28–35

Selected content from *Electronics Letters*. An invited Insight Letter looking at the field of 3D-HDTV

ISSN 2041-5923

Model behaviour for 3D-HDTV

D. Wood

EBU, Technical Department, Rt. Ancienne 17a, Grand Saconnex 1218, Switzerland
E-mail: wood@ebu.ch

Abstract: 3D occupies the unique position of being both the latest and the oldest technology in digital television. After short periods in the 1920s, 1950s and 1980s, it is set today to provide the next wave of both digital broadcast television and packaged media. Tools to help the standardisation and choice of 3D television are outlined. Today, only 'plano-stereoscopic' television systems are practical, but they are the first step in a 3D television evolution. This Letter suggests what the evolutionary path may be. Then a simple way to understand the inter-relationship between what the camera sees and what the viewer perceives for first-generation 3D is presented. Finally, a set of conditions that can be used for quality evaluation of first-generation 3D television is offered.

1 Introduction

3D has been subject to cyclic development, and enthusiasm, for many decades. The current interest in 3D television arises because digital processing now makes it easier to make, align and deliver a simple technical form of 3D, and high definition TV (HDTV), rather than standard definition TV (SDTV), allows higher quality 3D than previously possible. Research into more sophisticated 3D systems also continues. There is much to do in developing, evaluating and using 3D-HDTV. This work calls for 'models' that describe the contexts of the technology of 3D-HDTV. The models give those working in 3D television systems a common language to discuss alternatives and arrangements. This Letter outlines elements of three models that seem most valuable.

The first model, developed by the author and used by the ITU [1, 2], identifies the potential evolution of 3D over time, which will develop in a series of 'levels' and 'generations'. The second model, a development of work done for the cinema [3], is a transmission-system-independent, end-to-end, simple model for the first-generation 3D-HDTV ('S3D' HDTV, where the 'S' refers to stereoscopic). This should help us to understand the effects of changes of 3D operating parameters on the TV viewer experience. The third model is a reference model for evaluating alternative first-generation systems. When used as part of a complete evaluation methodology, this should help quantify quality differences in systems.

They are thus intended to help workers locate 3D-HDTV technologies in an evolution space, in a functional space and in a capability space.

2 3D system evolution model

Over the coming years, with advances in technology, 3D systems will provide a viewing experience that matches, ever more closely, our natural vision.

- There will be a long-term evolution of 3D television that begins with a 'first-generation' or plano-stereoscopic television (two-channel systems, L and R, termed S3D, see Fig. 1), which will progress, over what may be 6–12 years, into a second-generation multiview system (M3D); and, after a similar period, to third- and fourth-generation systems that allow more comfortable perception of depth. Each generation has fewer limitations than the last and comes closer to 'natural vision'.

- A characteristic of the first generation will be the predominant use of viewing glasses. The second generation will be glasses-free (auto-stereoscopic) and allow multiple pairs of views so that different parallax is seen with head movement. The third will provide a spatially quantised 'object wave', and the fourth a continuous object wave – in a sense natural vision. The object wave is the total lightwave that enters a given area, and which the eyes 'sample'. The 'hologram' is a simple form of object wave recording. Future research will provide more sophisticated

Figure 1 *Plano-stereoscopic images display left and right eye pictures on same display screen*

Glasses are used to direct correct picture to correct eye, to create impression of volume [Photo: DW]

ways to record the amplitude, phase and frequency of the object wave.

- There will be shorter-term development of S3D-HDTV within the first generation, with a series of alternative technologies, which fall into four levels (Table 1).

(i) The first, level 1 S3D-HDTV, is the ‘colour anaglyph’ process that uses different colour casts (largely of the luminance signal) for each of the L and R signals. It calls for no new equipment by the viewer, but has modest quality results.

(ii) The second level is the transmission of the L and R arranged in a spatial multiplex to fit a 2D-HDTV frame. This allows existing HDTV set top units to be used with new displays, but there are constraints on the resolution of the L and R pictures compared to normal HDTV pictures. The large screen display viewer will use glasses to select the appropriate temporally sequential L and R pictures. This is the form being used for the first 3D television broadcasting services. Small screens for handhelds will not use glasses, but special screens that direct adjacent left and right eye columns to the correct eye.

(iii) The third and fourth levels are the transmission of the L and R pictures in a way which allows full resolution L and R, and may provide features such as an embedded compatible 2D version of the picture, or depth adjustment in the display. These approaches are called level 3 and level 4, depending on the backward compatibility approach (level 2 or 2D-HDTV).

3 S3D-HDTV end-to-end delivery model

One way to describe the end-to-end delivery model of an S3D-HDTV system is as a transformation of objects from ‘real space’ into a ‘perceived space’, which the viewer mentally creates. The mental process is ‘stereopsis’, and what we create is the ‘cyclopean’ image – our perception is of objects, seen apparently from a point between our two eyes, with solidity. Assuming transparency of resolution,

Table 1 Four generations of 3D systems, with four levels of development for first generation

	First-generation 3D plano-stereoscopic (S3D HDTV)	Second-generation 3D multiview auto-stereoscopic (glasses free) display (M3D)	Third-generation 3D integral TV (quantised object wave) auto-stereoscopic display	Fourth-generation 3D continuous object wave recording auto-stereoscopic display
Level 4 new display and new set top box (similar approach to Blu-ray)	HDTV Service compatible (2D receivers decode a 2D version). 2D-HDTV + enhancement	–	–	–
Level 3 new display and new set top box	FC compatible (a signal ‘tops up’ the resolution of level 2). FC + enhancement	–	–	–
Level 2 new display but no new set top box (DVB phase 1, first 3DTV broadcasts)	HDTV frame compatible (3D signal appears as HDTV signal) L and R arranged in spatial multiplex	–	–	–
Level 1 no new display or set top box	Colour anaglyph. L and R have colour casts of complementary colours, such as red and cyan	–	–	–

noise and colourimetry, to create a simple model of the 3D-HDTV process, the relationship between objects in real space and in perceived space, S3D-HDTV may be considered as a function

$$D = F(d)$$

where D is the distance from the viewer to the perceived object, and 'd' is the distance from the camera to the real object.

If we can deduce the function 'F', we will have an end-to-end model of the S3D-HDTV delivery. The full mathematical analysis will be included in a future publication. (The work will be given in a future edition of the EBU Technical Review, which can be accessed via tech.ebu.ch.)

If we consider a perfectly transparent transmission or broadcast, it can be shown that

$$F = \frac{F1\,\text{(factors associated with the viewer's location in relation to the screen)}}{F2\,\text{(factors associated with the cameras' configuration, and the subsequent processing of the signals from them)}}$$

From this we can deduce the following simple rules:

1. The position in the 'z' or depth plane of the perceived object seen is affected linearly by the distance of the viewer from the screen. For example, if the viewer moves to be twice as far from the screen (perpendicular to the screen), the same object will have twice the depth. Thus, forward pointing arms of a given length will be perceived to be twice as long at twice the distance from the screen.

2. The position in the 'z' or depth plane of the perceived object is inversely proportional to the focal length of the taking cameras, all other things being equal. Assuming no processing, twice the focal length lens on the taking cameras will mean half the distance from the viewer to the perceived object.

3. All other things being equal, the position of the perceived object for the viewer is an inverse function of the spacing of the stereo cameras. A camera spacing that is wider than eye-spacing has other implications and gives the perceived objects a smaller appearance (the 'giant's eye view' phenomenon). The reverse applies for doser camera spacing (babies' eye view).

4. The two stereo cameras may be mounted to point directly forward, or they can be pointed in towards objects in the scene. This pointing inward is termed 'toe-in'. $F2$ would be affected by any 'toe-in' used for the camera. More toe-in will move objects forward in the perceived picture – but has other implications too, because the makeup of the background scene will depend on the degree of toe-in, and the left and right backgrounds will match each other less well with greater toe-in.

5. 3DTV systems need to be designed on the basis of assumptions about the shooting and viewing conditions

4 Reference model viewing environment

The quality potential of 2D television systems is evaluated using reference viewing conditions. A 'design viewing distance' is assigned to each category of television systems (HDTV, SDTV, extended definition TV (EDTV), and limited defnition TV (LDTV)). This is the distance at which the eye is considered 'saturated' with picture detail. Saturation is the point where the addition of more detail in the picture does not increase the perceived quality. It is expressed as multiples of picture heights. It is used for subjective evaluations of alternative 2D television systems. It is stringent, but it is equivalent to test driving a car by going 'up-hill'.

For analogue and digital HDTV systems, the design viewing distance is taken to be 3H (three times picture height). For digital SDTV (ITU Rec. 601, sometimes termed EDTV) it is 4H and for analogue SDTV (PAL, SECAM, NTSC) it is 6H. Even closer design viewing distances are assumed for 'ultra-high definition television'.

A model is now needed for S3D-HDTV, to allow repeatable evaluations of 3D technologies, and eventually to choose between systems, and to optimise them.

In the 3D-HDTV environment there are more elements than the usual 2D quality factors (such as colourimetry, noise, and spatio-temporal resolution) to consider. We need, *inter alia*, the following:

1. The geometrical congruency of the perceived image. This is the relationship between width/depth of real objects and those of the perceived objects. Programme makers are always free to arrange this for creative effect in a 3D programme, but the reference condition should arguably provide the viewer with a match to the scene in front of the camera – 'reality'. The viewing position of geometrical congruency is termed the orthostereoscopic condition. At this distance, the angle of view of the viewer matches the taking angle of the camera shooting the scene. In this position, everything in the scene looks 'right'. If we are closer to the screen than this, position depth is compressed compared to reality, and if we are further, depth is elongated.

2. In addition to horizontal and vertical resolution and congruency, achievable depth detail resolution affects quality of perception. This can be a variable for evaluation.

3. In addition, we need to consider the absolute physical distance of the viewer from the screen. S3D-HDTV systems can engender eye discomfort, and this can be associated with the combination of eye focusing (accommodation) and pointing of the eyes (convergence). In an S3D system, the eye must always focus on the screen to get the sharpest image, yet the disparity between the left and right eye signals, and the eyes' convergence, may be telling the eye to focus where the perceived object is. This is termed the potential accommodation–convergence conflict. This cannot be avoided, but it can be minimised if there is careful control in production grammar for the position of perceived objects. In addition, if the viewer-screen distance is greater than the eye's hyperfocal distance (HFD), we are more likely to cope without eye strain when accommodation and convergence do not match, because eye focus adjustments are not needed when the eye is focused at the HFD. The HFD of the eye (or any camera lens) is the point of focus where further focusing is not needed for sharp images. The normal human eye HFD is about three metres (3000 mm).

It can also be shown that the orthostereoscopic distance is the product of the ratio of the camera sensor size and the screen size (or M), and the focal length of the camera lens f_c

$$D\,(\text{orthostereoscopic viewing distance}) = M\,(\text{magnification factor}) \times f_c\,(\text{focal length of camera lens})$$

As an example, for a TV screen of 50 inch diagonal, and a 3DTV camera with a 1/2 inch diagonal camera sensor (as typical but not necessarily always the case), the magnification factor is 100. If we set the orthostereoscopic viewing distance at 3000 mm (the HFD), the programme needs to be shot with a 30 mm lens.

We can note the following:

1. The 3000 mm HFD viewing distance for a 50 inch display is about 5H (4.8H). To achieve a 3H viewing distance, which is assumed to be the approximate point of detail saturation in HDTV, would call for a screen size of about 80 inches. 80 inch displays are not typical of S3D-HDTV home installations today, and are not practical today for evaluations. We could also note that frame compatible systems provide less resolution than 'normal' HDTV.

2. Having a larger screen and greater viewing angle would contribute to the viewer's sense of immersion, and this would be a definite advantage for 3D viewing, but unless the absolute distance of the viewer to the screen is maintained at least at the HFD, the propensity for eye fatigue increases, which could bias evaluations.

3. The depth resolution available in a S3D-HDTV picture is technically related to the number of distinguishable horizontal picture elements possible in the on-screen 'infinity separation distance'. In simple terms, the more usable horizontal resolution in the L and R pictures, the more depth planes are possible in the 3D picture.

4. The overall impression of quality of a 3D picture and eye comfort will be influenced by many elements including impairments to depth cues and the registration of the L and R pictures.

The choice of a reference viewing environment for S3D-HDTV is necessarily a compromise, but bearing in mind the primary need to minimise potential eye discomfort, the best (or 'least worst') reference viewing environment today is probably the use of a 50 inch display and a 5H viewing distance, with camera lens as close to 30 mm as possible. Results will need to be interpreted bearing in mind that HDTV evaluation results are normally for 3H.

Evaluations may use adapted standard methods such as the EBU I, DSCQS, or EBU II method [4], though work remains to be done to establish full 3D methodology. There are probably three key quality factors to evaluate for S3D-HDTV. They are as follows:

- naturalness of images/degree of impairment and distortion in stereoscopic space

- viewing comfort

- visual fatigue.

5 Conclusions

This Letter has introduced three models for 3D-HDTV. The first locates 3D technology in evolutionary space. The second is an end-to-end model of the S3D-HDTV system. The third is a viewing reference model. They should provide a common language for the development and use of S3D-HDTV.

Will 3D's history of boom and bust be repeated this time, or will it become a long-term major part of television? The answer lies less in the hands of engineers and scientists than it did for colour TV or HDTV. This time, we will also sink or swim by the care with which 3D programmes are made. If programme makers avoid the visual grammar that leads to eye discomfort, viewers will be able to watch and enjoy 3D for sustained periods. If they do not, the 'water may be poisoned', and 3DTV may go back onto the shelf. If ever 3D television had a chance to succeed, because of the benefits of digitisation and HDTV, this is the time. But even if first-generation S3D does not succeed, the search for more natural viewing experiences will never stop, and eventually we will have ‘natural vision’ in our homes.

6 References

[1] ITU-R 6/177 New Report: ‘Features of three dimensional television (3DTV) video systems for broadcasting’, November 2009

[2] Wood, D.: ‘The truth about stereoscopic television’, *The Best of IET and IBC*, 2009, **1**, pp. 45–51

[3] SPOTTISWOODE R., SPOTTISWOODE N.L., SMITH C.: ‘Basic principles of three dimensional film’, *J. Soc. Motion Pict. Telev. Eng.*, 1952, **52**, pp. 249–285

[4] HOFFMANN H., WOOD D., ITEGAKI T.: ‘Physcho-physical methods of television picture quality evaluation’, *Electron. Lett.*, **43**, (4), pp. 212–213

Selected content from *Electronics Letters*

ISSN 2041-5923

Wireless digital data transmission at 300 GHz

C. Jastrow[1] *S. Priebe*[2] *B. Spitschan*[2] *J. Hartmann*[3]
M. Jacob[2] *T. Kürner*[2] *T. Schrader*[1] *T. Kleine-Ostmann*[1]

[1]*Physikalisch-Technische Bundesanstalt, Bundesallee 100, Braunschweig 38116, Germany*
[2]*Institut für Nachrichtentechnik, Technische Universität Braunschweig, Schleinitzstraße 22, Braunschweig 38106, Germany*
[3]*Rohde & Schwarz Vertriebs-GmbH, Vierenkamp 6, Hamburg 22453, Germany*
E-mail: christian.jastrow@ptb.de

Abstract: Recently, analogue video signal transmission at 300 GHz has been demonstrated using a versatile Schottky mixer based measurement system designed for terahertz communication channel modelling and propagation studies. In this reported work, digital signal transmission at 300 GHz using this system is demonstrated and analysed. The performance of the digital transmission setup is characterised with respect to phase noise and modulation errors. For demonstration, high data rate digital video signals have been transmitted over a distance of up to 52 m.

1 Introduction

The exponential growth of wireless data rates seen over the last thirty years, and the continuously increasing demand for unoccupied bandwidth, will lead to the extension of communication systems to higher frequencies in the lower terahertz (THz) range [1]. Owing to the existence of a suitable atmospheric transmission window such systems could operate at 300 GHz. The design of future digital THz communication systems will require channel characterisation with regard to path loss, phase noise, modulation and coding analysis at these frequencies, which has not been done yet.

In this Letter we present a setup for digital signal transmission based on a 300 GHz transmission system designed for channel measurements and propagation studies [2]. We discuss its phase noise and examine the performance of digital video broadcasting (DVB) transmission with special regard to link quality. Furthermore, we demonstrate the potential of digital signal transmission at 300 GHz by establishing a 52 m indoor link for 1080p full scale high definition television data which is a remarkable advance compared to the previous analogue video transmission experiment [2].

2 Setup

The 300 GHz transmission setup consists of autarkic transmitter and receiver units based on Schottky mixers, which are used to convert baseband signals with frequencies between 0 and 10 GHz and a maximum power of -3 dBm up to approximately 300 GHz (Fig. 1). The subharmonic mixers are pumped by frequency multiplier chains driven by two low phase noise signal generators used as tunable local oscillators. Transmitter and receiver are identical in construction aside from different frequencies used as input to the local oscillator chains (16.40 GHz at the transmitter against 16.38 GHz at the receiver) for noise suppression, resulting in an intermediate frequency of 360 MHz at the receiver output of the super-heterodyne system. A Rohde & Schwarz SFE broadcast tester directly connected to the mixer of the transmitter is used to generate DVB-T and DVB-S2 data streams of scalable frequency and power as input signals. On the receiver side either a Rohde & Schwarz ETL TV analyser for measurements on DVB-T signals or a DVB-S2 set-top-box is connected to the output to demonstrate high quality data transmission. To exploit the full dynamic range a transmission distance of 10 cm for the phase noise measurements and of 10 to 70 cm for the modulation analysis was chosen whereas two polyethylene lenses providing additional directional gain were used for the DVB-S2 demonstration to overcome high path losses.

3 Phase noise

To estimate the transmission performance and find suitable modulation schemes for the digital transmission link the

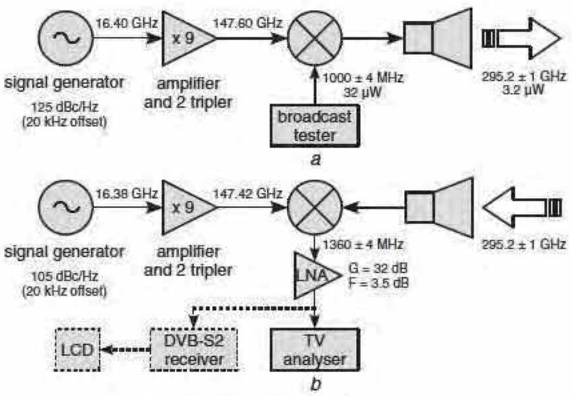
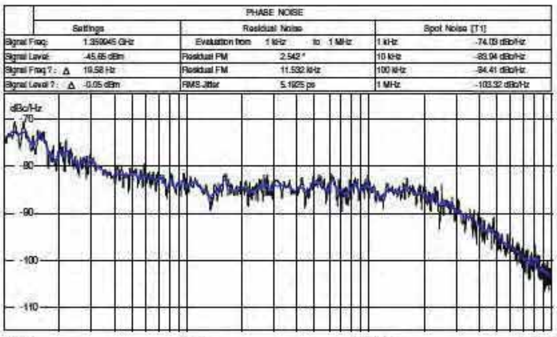


Figure 1 *Block diagram of digital transmission setup*

a Transmitter

b Receiver

phase noise performance of the 300 GHz transmission setup has to be evaluated first. The measurements were carried out using a spectrum analyser and a sinusoidal input signal of 1 GHz with a power of -5 dBm. The signal originating from a low phase noise signal generator (-130 dBc Hz^{-1}, 20 kHz offset) attached directly to the input of the transmitter was chosen to have the same frequency as the digital transmission. Even though the two local oscillators feature low single-sideband phase noise their contribution to the overall noise performance is predominant and constitutes the limiting factor when using high-order digital modulation schemes. This is because frequency multiplication inevitably leads to an increase of the phase noise level. Furthermore, the phase noise of both local oscillators is convolved into the intermediate frequency range, decreasing the signal-to-noise ratio at the receiver side substantially. Leading to inter-carrier interference, phase noise may heavily deteriorate the performance of modern modulation techniques such as OFDM (orthogonal frequency-division multiplex) [3]. Fig. 2 shows the noise floor of the received upper sideband signal evaluated between 1 kHz and 1 MHz. The phase noise level is below -73 dBc Hz^{-1} between 1 and 3 kHz and well below -80 dBc Hz^{-1} above 3 kHz. This noise performance is considerably good for a multiplier chain based transmission setup, being suitable for excellent transmission quality.



PHASE NOISE					
Settings		Residual Noise		Spot Noise [T1]	
Signal Freq:	1.359945 GHz	Evaluation from 1 kHz to 1 MHz		1 kHz	-74.03 dBc/Hz
Signal Level:	-45.65 dBm	Residual PM	2.542 °	10 kHz	-83.94 dBc/Hz
Signal Freq ? : Δ	19.58 Hz	Residual FM	11.532 kHz	100 kHz	-94.41 dBc/Hz
Signal Level ? : Δ	-0.05 dBm	RMS Jitter	5.1925 ps	1 MHz	-103.32 dBc/Hz

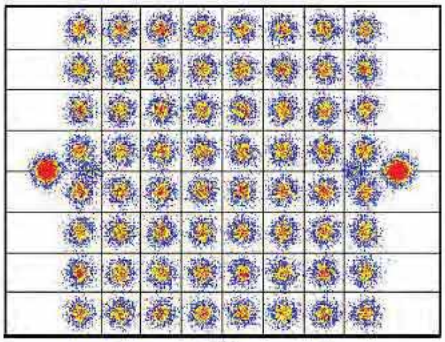
Figure 2 *Measured phase noise of 300 GHz transmission setup using baseband input signal of 1 GHz with input power of -5 dBm*

4 Digital transmission performance

Because future THz communication systems will undoubtedly use digital modulation schemes for data transmission, the influence of channel characteristics at 300 GHz and local oscillator phase noise on digital modulation has to be investigated. The broadcast tester generates a DVB-T signal with a centre frequency of 1 GHz and a channel bandwidth of 8 MHz. Owing to the high crest factor and a maximum allowable mixer input power of -3 dBm the signal power was limited to -15 dBm (-3 dBm peak envelop power) at the input of the transmitter resulting in a radiated signal power of 3.2 μW. To overcome high path losses a low noise amplifier was used to amplify the received signal to a level strong enough for the TV analyser. For transmission distances of 10, 30, 50 and 70 cm we analysed signal power, carrier-to-noise ratio (C/N), bit error rate (BER) and error vector magnitude (EVM), ensuring a signal integrity sufficient for proper demodulation of the DVB-T stream. Because of the excellent noise performance, the highest possible transmission mode designated for DVB-T was chosen. In detail, this mode uses a 64-QAM with a Viterbi forward error correction (FEC) code rate of 7/8 and a guard interval of 1/32, resulting in an MPEG transport stream bit rate of 31.668 Mbit/s in an 8 MHz channel. For this mode a C/N of 20.2 dB is necessary to ensure a quasi-error-free transmission in an AWGN (additive white Gaussian noise) channel [4]. For the multi-carrier OFDM system 1512 active subcarriers were chosen (2k mode) since both 4 and 8k modes exhibited temporary drop-outs and the demodulation process failed. We assume that inter-carrier interference caused by system phase noise is responsible for the drop-outs since the higher modes use a smaller carrier spacing, being more sensitive to inter-carrier interference [3]. The results of the measurements together with the calculated free space losses are shown in Fig. 3*a*. Concerning the BER, one can see that the critical limit of 2.0×10^{-4} before Reed Solomon error correction for a quasi-error-free transmission is not exceeded up to 70 cm. Nevertheless the decoding of the MPEG stream was barely successful at this distance. To leave a mark of the good transmission performance the constellation diagram of the received 64-QAM DVB-T signal is shown in Fig. 3*b* for a transmission distance of 30 cm with a BER of 2.6×10^{-8}. As typical for OFDM systems, phase noise shows up by the ideal signal states being expanded to form circular clouds centred at the nominal constellation points. However, these clusters are small enough not to exceed the demodulator's decision thresholds and to ensure a quasi-error-free transmission up to 70 cm.

Transmission distance (cm)	Free space loss (dB)	Signal power at analyser input (dBm)	C/N (dB)	BER before Reed Solomon	EVM (rms)
10	62.0	-36.4	35.4	$< 10^{-9}$	3.41%
30	71.5	-45.4	26.9	2.6×10^{-8}	4.40%
50	76.0	-50.4	21.5	1.2×10^{-4}	6.15%
70	78.9	-52.0	20.4	3.4×10^{-3}	6.89%

a



b

Figure 3 *Modulation analysis*

a Measured modulation parameters of received 64-QAM DVB-T signal in 2k mode with centre frequency of 1 GHz and input power of −15 dBm indicating digital signal quality for different distances

b Constellation diagram for transmission distance of 30 cm

5 Demonstration of transmission capabilities

To demonstrate the capabilities of the 300 GHz transmission system we chose a DVB-S2 signal containing a 1080p full scale high definition television data stream, also being generated by the broadcast tester SFE. A commercial DVB-S2 set-top-box was used to demodulate the transmitted signal, which was subsequently displayed on an off-the-shelf LCD TV set. For the transmission we employed the commonly used 8-PSK modulation, a symbol rate of 32.017 MS/s and a roll-off factor of 0.15, achieving a gross bit rate of 96 MBit/s in a 36.8 MHz-wide channel. With an FEC code rate of 9/10 an MPEG-2 net bit rate of 85.78 MBit/s could be attained [5]. With DVB-S2 being a single-carrier system with a crest factor lower than in multi-carrier systems, the input power applied to the mixer was increased to −10 dBm. By using the two polyethylene lenses resulting in an overall antenna gain of 40 dB per unit the transmission range could be increased up to 52 m. A quasi-error-free transmission link could be established even when the input power had been reduced by 8 dB to a radiated signal power of 1.6 μW. Apart from the collimation of the 300 GHz radiation into a nearly parallel beam, this is due to the powerful FEC system based on an highly-efficient inner LDPC code concatenated with an outer BCH code.

6 Conclusion

We have shown first measurements regarding modulation analysis and BER at 300 GHz in a 64-QAM modulated OFDM channel. Additionally, by transmitting a 96 MBit/s DVB-S2 signal over a distance of 52 m, we undoubtedly showed the feasibility of high data rate communication links in the lower THz frequency range using high-order modulation schemes with a suitable forward error correction. The results indicate that the system limits have not been reached, yet. However, further investigations at ultra-high data rates will require a suitable unidirectional data source.

7 Acknowledgments

We wish to thank P. Schlegel and A. Gudat, both with the Institut für Nachrichtentechnik at TU Braunschweig, for the supply of measurement equipment and participation in preliminary experiments, as well as for helpful discussions.

8 References

[1] PIESIEWICZ R., KLEINE-OSTMANN T., KRUMBHOLZ N., MITTLEMAN D., KOCH M., SCHOEBEL J., KÜRNER T.: 'Short-range ultra broadband terahertz communications: concept and perspectives', *IEEE Antennas Propag. Mag.*, 2007, **49**, (6), pp. 24–39

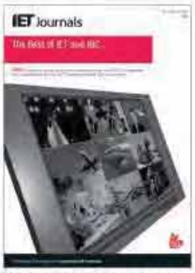
[2] JASTROW C., MÜNTER K., PIESIEWICZ R., KÜRNER T., KOCH M., KLEINE-OSTMANN T.: '300 GHz transmission system', *Electron. Lett.*, 2008, **44**, (3), pp. 213–214

[3] PETROVIC D., RAVE W., FETTWEIS G.: 'Intercarrier interference due to phase noise in OFDM – estimation and suppression'. Vehicular Technology Conf., Los Angeles, CA, USA, , vol. 3, September 2004, pp. 2191–2195

[4] DVB-T specification ETSI EN 300 744 v1.6.1, 'Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television', p. 40

[5] DVB-S2 specification ETSI EN 302 307 v1.1.2, 'Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications'

Selected content from *Electronics Letters*



ISSN 2041-5923

Formel-Kapitel 1 Abschnitt 1A software defined radio realisation of DVB-T2 receiver

A. Vie&szlig;mann[1] A. Waadt[1] C. Spiegel[1] C. Kocks[1] A. Burnic[1] P. Jung[1] G.H. Bruck[1] J. Kim[2] J. Lim[2] H.W. Lee[2]

[1]Universität Duisburg-Essen, Lehrstuhl für Kommunikations technik, Oststrasse 99, Duisburg D-47057, Germany
[2]Samsung Electronics, Global Standards and Research Lab.
E-mail: Peter.Jung@KommunikationsTechnik.org

Abstract: DVB-T2, the revision of DVB-T (Terrestrial Digital Video Broadcasting), was recently finalised, being targeted to high definition television (HDTV). To pave the way to commercialisation, an appropriate implementation concept and its corresponding validation are of utmost importance. Without a doubt, the most challenging requirements introduced by the DVB-T2 specification are a fast Fourier transform size of up to 32 k samples, 256-QAM (quadrature amplitude modulation) and low density parity check coding with a block size of 64 k bits. In this Letter, a software defined radio realisation concept is presented, comprising a combined digital signal processor and field programmable gate array solution, which is tailored to meet these requirements.

1 Introduction

DVB-T (digital video broadcasting, terrestrial) has established itself as the leading specification for digital terrestrial TV broadcasting. Since DVB-T was initially specified, there have been substantial developments in both the modulation technology available and the economics of the transmission chain. These recent results have paved the way towards a renovation of DVB-T, enabling new ways of modulating and error-protecting the broadcast stream and, as a consequence, an increased number of programmes. Meanwhile, the standardisation of the second generation DVB-T2 has been completed [1] and first realisations of new broadcasting networks are expected to appear within the coming months.

With the decision to apply higher-order modulation (HOM) up to 256-QAM (quadrature amplitude modulation), forward error correction coding based on low density parity check (LDPC) codes with codeword lengths of up to 64 k bits, and orthogonal frequency division multiplexing (OFDM) using a gross number of up to 32 k points [1], DVB-T2 poses critical real-time requirements on the implementation of receivers. In addition, the receivers are required to be flexible to allow the defined variations of e.g. the modulation scheme, the channel coding and the number of OFDM subcarriers [1]. It is therefore recommendable to base the receiver designs on the software defined radio (SDR) paradigm [2]. However, state-of-the-art digital signal processors (DSPS) are not powerful enough to provide a full DVB-T2 receiver implementation. Therefore, alternative approaches are required which makes the implementation of well assorted high-speed connections between the different receiver components a mandatory prerequisite. In this Letter, the authors present an SDR based DVB-T2 receiver concept which was implemented and successfully tested at the Lehrstuhl für KommunikationsTechnik.

2 SDR based DVB-T2 receiver concept

The SDR based DVB-T2 receiver developed by the authors comprises a radio frequency (RF) front end, which allows the processing of signals in various frequency bands; a mixed signal board which allows the analogue-to-digital-conversion of the down-converted analogue received signals and the distribution of the digital signals to the various digital processing entities through a small XILINX SPARTAN3 field programmable gate array (FPGA); a

DSP board with a powerful Texas Instruments TMS320C6455 DSP and an FPGA board with a large XILINX VIRTEX5-LX110 FPGA. The DSP and the large FPGA provide the digital signal processing capabilities required by the DVB-T2 standard.

The DSP is the heart of the DVB-T2 receiver, controlling and reconfiguring all further hardware components. The large FPGA acts as a necessary and flexibly adjustable hardware accelerator, providing real-time error control decoding capabilities. The DSP is programmed in C++ language using the Code Composer Studio, and all FPGAs are programmed in Verilog hardware description language (HDL) using the XILINX ISE development suite.

To facilitate a high speed interconnection between the DSP and the large FPGA, the authors have developed a point-to-point variant, i.e. a simplified version, of the well-known peripheral component interconnect (PCI) bus. The error control decoded bit stream is the output of the SDR based DVB-T2 receiver developed by the authors. To facilitate a low-cost and at the same time standardised interface, the authors rely on a USB2.0 connection.

The RF received signals prevailing either in the VHF and the UHF bands are first processed by a Thomson DTT73200 digital terrestrial tuner, generating an intermediate frequency (IF) received signal at its output. To alleviate the impact of intermodulation distortions (IMDs) caused by I/Q imbalancing, the authors deploy a single heterodyne analogue receiver which feeds an Analog Devices AD6655 analogue-to-digital converter (ADC). In the ADC, the IF received signal is sampled and digitally downconverted. The digitised received signal output by the ADC is transferred via the small FPGA to the DSP for synchronisation and demodulation. The demodulated signal is then transferred to the large FPGA for LDPC decoding. Finally, the error control decoded bit stream is transferred to a quad-core host personal computer (PC) via a USB2.0 carrying out the source decoding and the video displaying via a high definition multimedia interface (HDMI) connection to a full HD display.

3 SDR based DVB-T2 receiver implementation

Fig. 1 is a photograph of the implemented SDR based DVB-T2 receiver. On the left-hand side, a compound of three stacked printed circuit boards (PCBs) can be seen. The top is the RF tuner board developed by the authors. This RF tuner board hosts the above-mentioned RF tuner. The RF tuner board is connected to the mixed signal board realised by the authors. This mixed signal board hosts two small FPGAs of which one is required for the DVB-T2 receiver. The mixed signal board is connected to the DSP board which contains the DSP operated at a clock frequency of 1.2 GHz. On the right-hand side of the photograph, the FPGA board carrying the large FPGA is located. This board is electrically connected to the mixed signal board via a PCI cable.

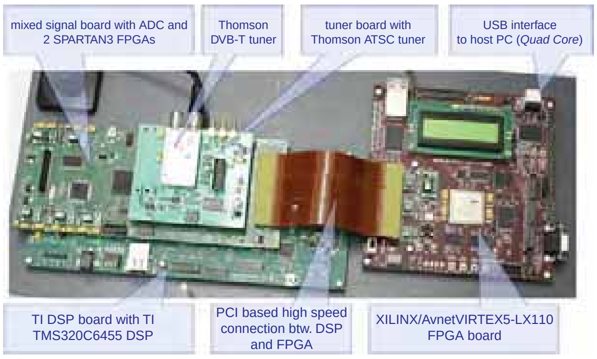

mixed signal board with ADC and 2 SPARTAN3 FPGAs
Thomson DVB-T tuner
tuner board with Thomson ATSC tuner
USB interface to host PC (*Quad Core*)
TI DSP board with TI TMS320C6455 DSP
PCI based high speed connection btw. DSP and FPGA
XILINX/AvnetVIRTEX5-LX110 FPGA board

Figure 1 *Photo of implemented SDR based DVB-T2 receiver*

4 Performance

Fig. 2 shows the obtained error performance of the demonstrator in the case of a transmission via a single path channel with additive white Gaussian noise (AWGN), assuming LDPC coding with a code rate of 3/5, a codeword length of 64 k bits, 32 k FFT and 256-QAM modulated data symbols transmitted over each subcarrier. This error performance was measured for benchmarking. In Fig. 2, both the bit error ratio (BER) P_{bit} as well as the block error ratio (BLER) P_{block} are shown against the signal-to-noise ratio (SNR) $10 \log_{10} (E_s/N_0)$. Both P_{bit} and P_{block} are determined at the output of the BCH decoder which follows the LDPC decoder. The LDPC decoding results were determined after 50 decoding iterations. Both LDPC and BCH decoders are soft-input decoders. It was found that P_{block} 10 ^{3} requires an SNR $10 \log_{10} (E_s/N_0)$ of less than 16.7 dB. The corresponding BER is approximately 10 ^{7} at the same SNR.

It was found that the analogue single heterodyne receiver allows a superb error vector magnitude (EVM) of less than 2%. The frequency offset of the RF tuner is lower than 6 kHz in the free-running mode. This frequency offset can be easily corrected by an automatic frequency correction (AFC) and synchronisation algorithm developed and implemented by the authors.

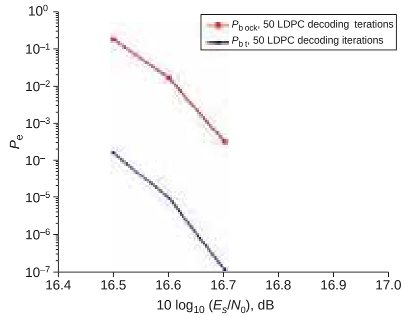


Figure 2 *Error performance*

The implemented SDR based DVB-T2 receiver was tested using an RF signal which was generated and transmitted through a Rohde & Schwarz AFQ/SMIQ06 combination. The test signal was based on a 256-QAM/32 k FFT variant of the DVB-T2 signal, often abbreviated as the PP7 setup. It contained a multiplexed version of three HDTV test video streams provided by the BBC. The gross information rate was 45 Mbit/s. Validation of the setup was successful.

5 Conclusions

The authors have described the hardware concept and setup of an SDR based DVB-T2 receiver. This receiver combines the digital processing power of a high performance DSP and a large FPGA with a single heterodyne analogue receiver concept. The authors believe that the successful demonstration of the setup will ease the way towards commercial and highly integrated implementations of future DVB-T2 receivers.

6 Acknowledgments

The authors thank F. Kirch and his colleagues at Rohde & Schwarz for their generous support of the project, providing an AFQ device for testing purposes, as well as their colleagues at Samsung Electronics and at the Lehrstuhl für Kommunikationstechnik of the University Duisburg-Essen for fruitful discussions.

7 References

[1] Digital Video Broadcasting (DVB); frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2). Draft ETSI EN 302 755 V1.1.1 (2008-04)

[2] JONDRAL F.K.: ‘Software-defined radio – basics and evolution to cognitive radio’, *EURASIP J. Wirel. Commun. Netw.*, 2005, **3**, pp. 275–283

Selected content from *Electronics Letters*

ISSN 2041-5923

Frame structure for DTTB uplink systems using novel training sequences

C. Zhang Z. Wang Z. Yang

*Tsinghua National Laboratory for Information Science and Technology, Department of Electronics Engineering,
Tsinghua University, Beijing 100084, People's Republic of China
E-mail: z_c@mail.tsinghua.edu.cn*

Abstract: A new frame structure based on time-domain synchronous orthogonal frequency division multiplexing for digital terrestrial television broadcasting (DTTB) uplink systems is proposed, whereby multicarrier pseudonoise (PNMC) training sequences are used as the frame header (FH). Sub-sequences of FH are inserted between signal frames as guard intervals. Simple multi-user channel estimation and equalisation can be applied and good performance is illustrated by simulations.

1 Introduction

The multi-user uplink access technique becomes important in digital terrestrial television broadcasting (DTTB) systems, especially for digital interactive television services. Recently, the European Telecommunications Standards Institute (ETSI) has launched the DVB-RCT standard [1], which uses orthogonal frequency division multiple access (OFDMA) as its uplink solution. In 2006, our time-domain synchronous orthogonal frequency division multiplexing (TDS-OFDM) proposal was adopted by the Chinese DTTB standard [2]. However, when the conventional TDS-OFDM technique is applied to multi-user frequency division multiple access (FDMA), time-domain pseudonoise (PN) sequences will bring severe multi-user interference (MUI). In this Letter, a novel frame structure is proposed to enable the usage of TDS-OFDM for multi-user uplink DTTB scenarios. Owing to its good performance and low complexity, it is a potential multi-user uplink solution for Chinese next generation DTTB systems.

2 Frame structure

Fig. 1 illustrates the proposed frame structure of one specific user for the multi-user uplink FDMA systems. The signal frame is the basic unit, which composes a data block and a guard interval. Several signal frames construct a super-frame. A training sequence $\boldsymbol{S}^i$ is placed in front of the super-frame, where i is the user index. It could be used for channel estimation and equalisation, as well as synchronisation. $\boldsymbol{S}^i$ is divided into two sub-sequences equally in the time-domain, $\boldsymbol{S}_0^i$ and $\boldsymbol{S}_1^i$, which are also inserted alternatively between data blocks as the guard intervals, as shown in Fig. 1. Assuming that the number of signal frames in one super-frame is K, for the kth signal frame $(k = 0, 1, \ldots, K-1)$, sub-sequence $\boldsymbol{S}_{\text{mod}(k,2)}^i$ is selected as the guard interval, where mod $(k, 2)$ is the modular arithmetic of 2. Normally K is even for the fixed training sequences. Otherwise, the order of $\boldsymbol{S}_0^i$ and $\boldsymbol{S}_1^i$ of the training sequences between adjacent super-frames should be exchanged when K is odd.

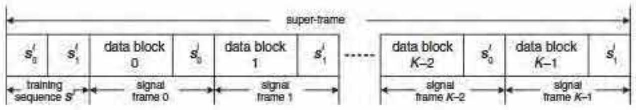


Figure 1 *Proposed frame structure*

3 System model

The specially selected multicarrier pseudonoise (PNMC) sequences are applied instead of the time-domain PN sequences used in the traditional TDS-OFDM systems. PNMC sequences are binary sequences in the frequency domain with length of M, $\boldsymbol{P} = [P_0, P_1, \ldots, P_{M-1}]^{\mathrm{T}}$. The property of constant modulus in the frequency domain can improve the accuracy of channel estimation with low complexity. PNMC is more suitable for FDMA without MUI.

Assuming that the total number of active users is W, the sum of the training sequences of all the users will make up a whole PNMC sequence. For the ith user, frequency-domain sub-sequence $\boldsymbol{S}^i$ is $\boldsymbol{S}^i \quad \boldsymbol{D}^i_M\boldsymbol{P}$, where $\boldsymbol{D}^i_M$ is an $M \times M$ sub-channel selection matrix. After M-point inverse discrete Fourier transformation (IDFT), the time-domain training sequence $\boldsymbol{S}^i$ with length M is generated. The synchronisation in time, power and frequency among different users for typical uplink scenarios can be achieved by the ranging procedure [1]. To ensure the orthogonality in the frequency-domain, $\boldsymbol{D}^i_M$ should be orthogonal between different users, which can be expressed as

$$\begin{cases} \sum_{i=0}^{W\quad 1} \boldsymbol{D}^i_M = \boldsymbol{I}_M \\ \boldsymbol{D}^i_M \times \boldsymbol{D}^j_M = \boldsymbol{O}_M \quad i, ; j \in [0, W-1],\ i \neq j \end{cases} \tag{1}$$

where $\boldsymbol{I}_M$ is an identify matrix with size of $M \times M$ and $\boldsymbol{O}_M$ is an $M \times M$ all-zero matrix.

OFDMA or single-carrier frequency division multiple access (SC-FDMA) can be applied to the data blocks. The total number of subcarriers is assumed to be N, then each user can use one subchannel including N/W subcarriers. The same subchannel and bandwidth must be occupied by data blocks and their corresponding training sequences.

The length of the guard interval is half of the training sequence, which means that the maximum multipath delay spread is $M/2$. The length relationship between guard intervals and training sequences is carefully selected owing to the property of DFT processing. Because the training sequences from different users are overlapped in the time-domain, only frequency-domain channel estimation can be used. For M-point DFT, when the multipath delay exceeds $M/2$, the results of DFT are under-sampled and the channel frequency response (CFR) after filtering and interpolating is not correct. Accordingly, the handled channel length is $M/2$ using M-point frequency-domain channel estimation. Based on the proposed architecture, the sub-optimum trade-off between maximum channel delay spread and spectrum efficiency is achieved.

4 Channel estimation and equalisation

The guard interval $\boldsymbol{S}^i_1$ of the last signal frame in the previous super-frame can be regarded as the cyclic prefix (CP) of the current training sequence $\boldsymbol{S}^i$. Therefore, frequency-domain channel estimation can be applied in the receiver. However, the received data blocks do not satisfy the cyclic convolution with the existence of multipath delay spread. The simple addition and subtraction can be used to reconstruct the CP-OFDM [3], as illustrated in Fig. 2.

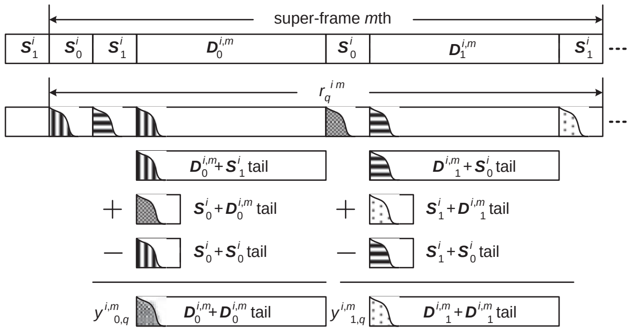


Figure 2 *CP-OFDM reconstruction of proposed system*

Assuming that the maximum multipath delay spread is L and $K \quad 2$, for the mth super-frame of the ith user, the two data blocks are denoted as $\boldsymbol{d}^{i,m}_0$ and $\boldsymbol{d}^{i,m}_1$, respectively, and the received super-frame is $\{r^{i,m}_q\}^{2M+2N\quad 1}_{q=0}$, where q is the sample point index. The reconstructed signal $\{y^{i,m}_{0,q}\}^{N\quad 1}_{q=0}$ and $\{y^{i,m}_{1,q}\}^{N\quad 1}_{q=0}$ can be expressed as

$$y^{i,m}_{0,q} = \begin{cases} r^{i,m}_{q+M} + r^{i,m}_{q+M+N} - r^{i,m}_q & 0 \le q < L \\ r^{i,m}_{q+M} & L \le q < N \end{cases}$$

$$y^{i,m}_{1,q} = \begin{cases} r^{i,m}_{q+3M/2+N} + r^{i,m}_{q+3M/2+2N} - r^{i,m}_{q+M/2} & 0 \le q < L \\ r^{i,m}_{q+3M/2+N} & L \le q < N \end{cases} \tag{2}$$

Because all signal frames of different users have the same property, the receiver can process the received signals once for all users simultaneously. The signal-to-noise ratio (SNR) loss due to the CP-OFDM reconstruction can be described as

$$SNR_{loss} = 10\log_{10}\left(\frac{N+2L}{N}\right) \tag{3}$$

5 Simulation results

The performance of the proposed system and DVB-RCT are estimated over static multipath fading channels, whereby ITU vehicular A channel (ITU A) [4] and DVB Ricean channel (DVB F1) [5] are selected. The system parameters are summarised in Table 1. The same concatenated Reed-Solomon encoding and convolutional encoding are used with code rate of 1/2. At the receiver side, the soft decision Viterbi decoder is used. Fig. 3 shows the average bit error ratio (BER) performance over all subchannels of the two systems. It can be seen from Fig. 3 that in ITU A channel about 5 dB gain can be achieved using the proposed system at BER $\quad 2 \times 10^{\quad 4}$, which is mainly due to the localised subcarrier allocation in the proposed system and the higher sampling rate in DVB-RCT. In DVB F1 channel, the performance of the proposed system is about 2 dB better than the DVB-RCT system at BER $\quad 2 \times 10^{\quad 4}$.

The actual bit rate per subcarrier of the proposed system according to Table 1 is 3.164 kbit/s, which is slightly

Table 1 System parameters

	Proposed system	DVB-RCT
Bandwidth	7.56 MHz	7.643 MHz
Sampling rate	7.56 MHz	9.143 MHz
Total subcarriers	2048	1711
Number of subcarriers for one subchannel	24	29
Number of subchannels	85	59
Constellation mode	QPSK	QPSK
Code rate	1/2	1/2
Number of data symbols in one burst	144 symbols	144 symbols
Number of OFDM symbols in one burst	6	6
Guard interval	1/8	1/8

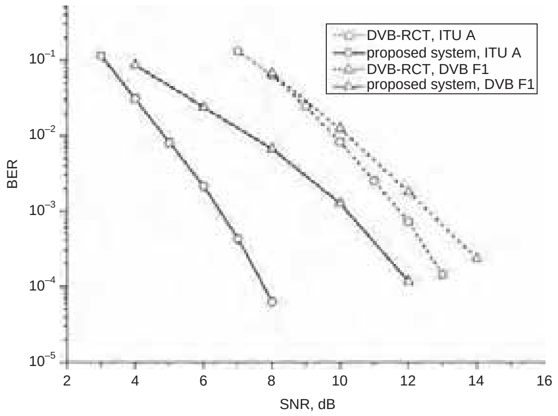


Figure 3 *Simulation results over multipath fading channels*

higher than 3.04 kbit/s of DVB-RCT. To improve the performance in the time-variant channels, one super-frame might contain only two signal frames. Therefore, the channel estimation can be updated every two signal frames, whereas six OFDM symbols are needed in the DVB-RCT system to obtain one complete CFR estimation. At this time, the payload bit rate per subcarrier is slightly decreased to 2.953 kbit/s.

6 Conclusions

A new frame structure using specific training sequences based on TDS-OFDM is proposed for multi-user uplink DTTB systems. Simulations show that better performance can be achieved under various multipath fading channels.

7 Acknowledgments

This work was supported by Tsinghua University Initiative Scientific Research Program and in part by the China National Science Foundation under grant 60902003.

8 References

[1] European Telecommunications Standards Institute (ETSI), 'Digital video broadcasting (DVB); Interaction channel for Digital Terrestrial Television (RCT) incorporating multiple access OFDM', EN 301 958, v1.1.1, ETSI, March 2002

[2] Chinese National Standard GB 20600-2006: 'Framing structure, channel coding and modulation for digital television terrestrial broadcasting system (in Chinese)', 18 August 2006

[3] FU J., WANG J., SONG J., PAN C.Y., YANG Z.X.: 'A simplified equalization method for dual PN-sequence padding TDS-OFDM systems', *IEEE Trans. Broadcast.*, 2008, **54**, (4), pp. 825–830

[4] ITU-R Recommendation M.1225, 'Guidelines for evaluation of radio transmission technologies for IMT-2000', 1997

[5] European Telecommunications Standards Institute (ETSI), 'Digital video broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television', EN 300 744, V 1.5.1, ETSI, November 2004