International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BT.2246
(10/2011)

The present state of ultra high definition television

BT Series

Broadcasting service
(television)

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at http://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management

Note: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.

REPORT ITU-R BT.2246

The present state of ultra high definition television

(2011)

Summary

This Report contains the results of the study on ultra-high definition television (UHDTV) and describes the present state of UHDTV. It addresses the baseband image format and the derivation of system parameter values in particular.

TABLE OF CONTENTS

1 Introduction

Television has built its history on the fundamental desire of human beings to extend their audio-visual senses spatially and temporally. HDTV is one of the great achievements of television. People in many countries are now enjoying the benefits of HDTV, and people in the rest of the world will soon benefit in the near future.

The definition of HDTV is described in Report ITU-R BT.801 The present state of high-definition television as follows:

> *A high-definition system is a system designed to allow viewing at about three times the picture height, such that the system is virtually, or nearly, transparent to the quality of portrayal that would have been perceived in the original scene or performance by a discerning viewer with normal visual acuity. Such factors include improved motion portrayal and improved perception of depth.*

The attainment of this goal is limited in some aspects; e.g. the field of view of HDTV is only 30 arc-degrees. Our natural desire to overcome such limitations has led us to the concept of UHDTV;

> UHDTV is a television application that is intended to provide viewers with an enhanced visual experience primarily by offering a wide field of view that virtually covers all of the human visual field with appropriate sizes of screens relevant to usage at home and in public places.

2 General considerations

2.1 Fundamental idea of UHDTV

UHDTV is a television application that will provide viewers with a better visual experience primarily by offering a wide field of view (FOV) which virtually covers all of the human visual field, while maintaining other features of HDTV or improving them. UHDTV could therefore be characterized as a TV system having a wide field of view supported by enhanced spatial resolution.

2.2 Classification of image systems

Figure 1 classifies the image systems that ITU-R deals with in terms of resolution and screen size. Resolution is one of the most important parameters in ITU-R Recommendations for television systems (image formats).Screen size has not been regarded as a parameter intrinsic to image formats.

The resolution of HDTV is described in Recommendation ITU-R BT.709 as 1920 × 1080.The resolution of LSDI is described in Recommendation ITU-R BT.1680, and the resolution for an expanded hierarchy of LSDI, namely 3840 × 2160 and 7680 × 4320, is described in Recommendation ITU-R BT.1769. The screen size of LSDI is described in Question ITU-R 15-2/6 as … *large screen presentation in ... theatres, halls, and other venues.*

The resolution of EHRI is described in Recommendation ITU-R BT.1201 as a simple integer multiple of 1920 × 1080. The specific values given in Report ITU-R BT.2042 are 1, 2, 3 and 4. The screen size of EHRI is not described anywhere in ITU-R documents since its basic idea is not related to the screen size.

From the conceptual viewpoint, it is natural to suppose that the resolution would exceed that of HDTV. It would also be natural to suppose that the screen size would be suitable for home UHDTV receivers. Nevertheless, as in the case of HDTV, the applications of UHDTV would probably not be restricted to home use. Therefore, it is reasonable to conclude that the screen size of UHDTV would not be limited to the range applicable to home use.

FIGURE 1

Classification of image systems
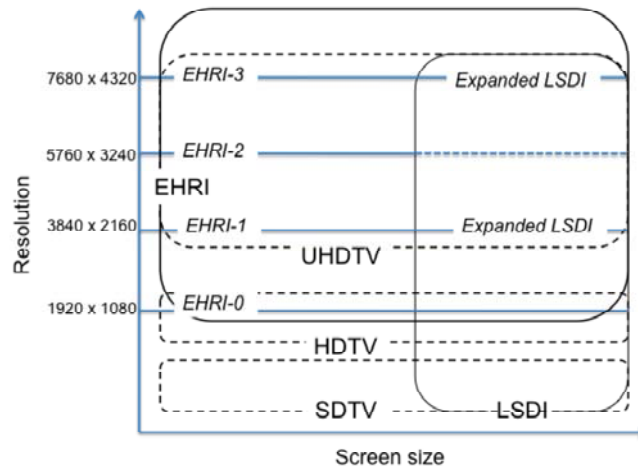
2.3 Application

2.3.1 Overview

Compared with current HDTV, the UHDTV application should bring considerably improved benefits to its viewers. Those benefits may include:

– stronger sensation of reality or presence;

– higher transparency to the real world;

– more information.

It may be presented in:

– living rooms;

– personal spaces in mobile and non-mobile environments;

– collective viewing locations such as theatres.

Each screen should be well suited to the particular form of usage.

Rec. ITU-R BT.1845, "Guidelines on metrics to be used when tailoring television programmes to broadcasting applications at various image quality levels, display sizes and aspect ratios", shows the relationship between the screen size and viewing distance given that the optimal viewing distance is one at which one pixel corresponds to the visual angle of one arc-minute. The optimal viewing distance is, for example, 100 cm when viewing an UHD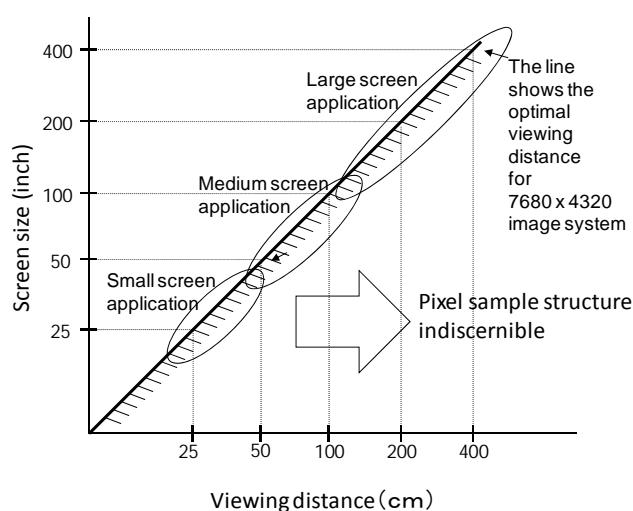TV image on a 100-inch screen. One pixel per one arc-minute is generally regarded to correspond to a visual acuity of 1.0. This means that the pixel sampling structure is not detectable when viewing under the conditions on the right and below the line in Fig. 2.

FIGURE 2

Screen size and viewing distance of UHDTV applications

A viewer typically watches a television while remaining stationary at around two to three meters from the display. Such a viewing style for UHDTV corresponds to a large screen application (Fig. 2) that may include theatrical environments as well as home environments. It will provide stronger sensation of reality and a stronger immersive feeling to viewers by offering a far wider field of view than current systems can offer.

The second category in Fig. 2 is a medium-size screen application with screen sizes from around 50 to 150 inches. The respective optimal viewing distances are 50 to 150 cm. The viewing style for this application may include not only the conventional one but also a new one in which viewers changes their viewing distance according to their preference or the content creator's intention. For example, fine art can be appreciated by standing as close as the painters painted them at. So too, with a UHDTV image of fine art.

The third category is a small-screen application. Similar applications for text data or still images have recently emerged and are known as electronic paper. The optimal viewing distance for a 20-inch screen is 20 cm, at which the HVS reaches the accommodation limit. The size of a 20-inch screen is close to A3, and a 7680 x 4320 pixel screen would have approximately 350 pixels per inch.

The use-cases described above are in accordance with our natural desire to extend our sensory limits. Some of them exceed the capabilities of conventional television. They should be studied as part of the development of future television applications.

The transmission paths to the presentation venue may vary depending on technical and economic reasons. They may include:

–   satellite radio transmission;

–   terrestrial radio transmission;

–   cable and network;

–   new wide-band networks, radio or cable, that can accommodate virtually lossless UHDTV signals;

–   package delivery.

In [Input, 9], a broadcaster expressed its view on the introduction of UHDTV as follows.

–   If we imagine a giant screen and a multichannel audio system with as many as 22.2 channels, the image and sound presentation arrangements required by the UHDTV are not likely to be broadly implemented or implementable in consumer homes. They might perhaps be implementable to some extent in a dedicated "TV room", which may more properly be called a "home-cinema room".

–   Both UHDTV systems, and particularly the 7 680 × 4 320 system will likely find their most successful application in LSDI, namely in the large-screen presentation of digital programmes for collective viewing under controlled conditions, in suitably equipped venues such as auditoria, cinema-like halls and similar places, on condition, of course, that programmes in UHDTV will be available in adequate number and quality to support such a broadcasting service.

2.3.2   Considerations of audience viewing

Worldwide television audiences have acquired certain viewing habits, which they might be reluctant to change unless the viewing experience is enhanced greatly.

A typical viewing habit is to watch television while sitting on the sofa in the living room. The TV screen is watched from a position perhaps 7 feet (2.1 meters) away. This viewing distance did not change when television transitioned from black and white to colour, or from SDTV to HDTV. Indeed, in the latter transition, the viewing distance remained roughly the same, while the TV screen became larger.

A 52" (132 cm) HDTV 1080" receiver today is viewed at a distance of about 7 feet (2.1M). Assuming that the home viewing habits will not change dramatically when transitioning from HDTV to 3 840 × 2 160 UHDTV, it can be expected that 100" (2.5 m) screens may well exist, in some instances such a large screen would take up a good part of the living room wall. This could be considered an enhanced viewing condition.

It cannot be excluded that consumers would be prepared to install such a large screen in their living room, but it seems unlikely that they will be prepared to install the 200" (5 m) screen that could be recommended for 7 680 × 4 320 UHDTV viewing.

It could be concluded that, from the point of view of image presentation, 3 840 × 2 160 UHDTV television broadcasting to the home may find acceptance with in-home television audiences, while it may be unlikely that 7 680 × 4 320 UHDTV will do so. The 7 680 × 4 320 UHDTV image system may instead find applications for television presentations to the public in theatres, in home theatres, auditoria, theme parks, and other public venues.

2.4 System model

Figure 3 depicts a possible system model for the applications described above.

FIGURE 3

Possible system model of UHDTV
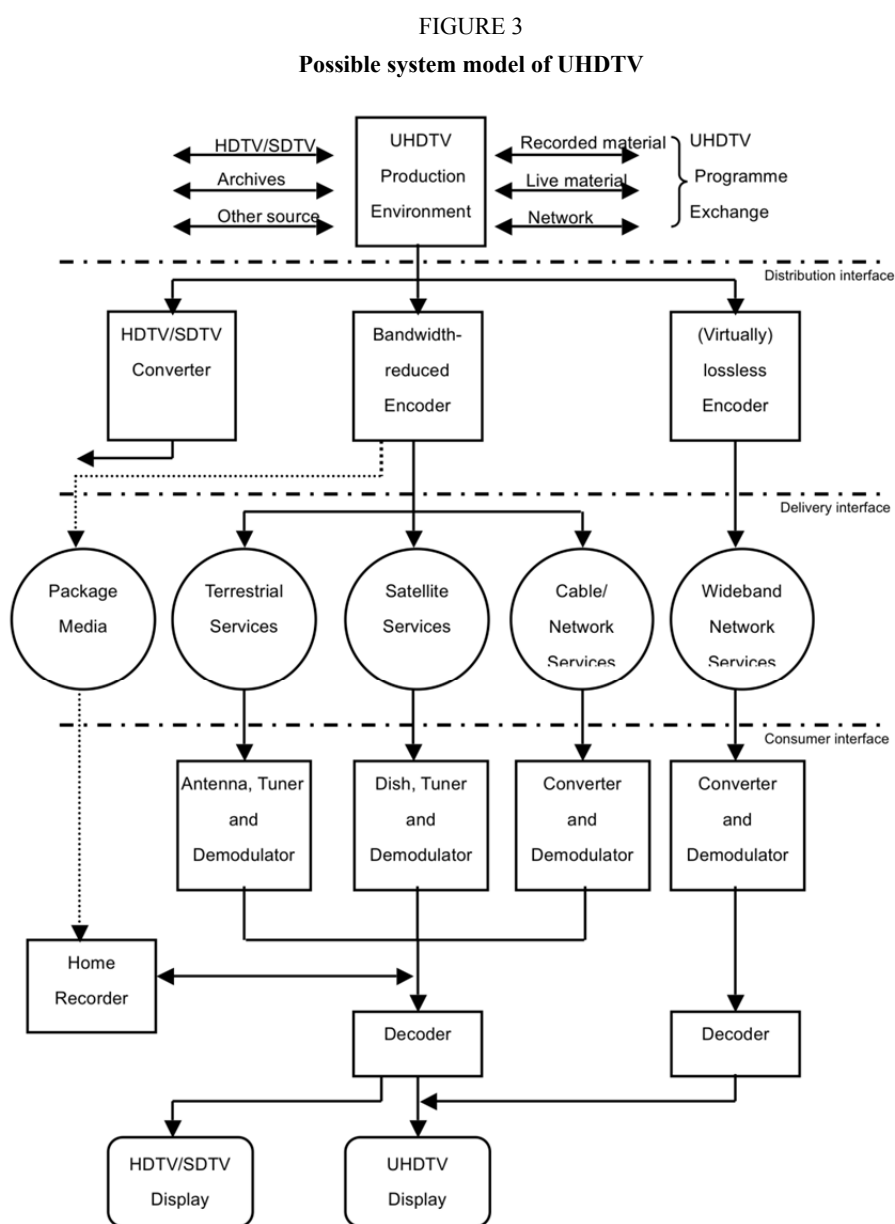
2.5 Viewing conditions

2.5.1 Field of view

Among the properties of current television pictures, field of view may be the one that is most inferior to the capabilities of the human visual system. Nevertheless, UHDTV does not intend to cover the whole surrounding field of view like that of virtual reality. The required value for a two-dimensional display to realize the applications described above should be explored.

2.5.2   Angular resolution

The "design viewing distance" gives the angular resolution (pixel/unit visual angle). It has tacitly been taken to be one pixel/arc-minute. Recommendation ITU-R BT.1127 "Relative quality requirements of television broadcast systems" describes the design viewing distance as the relative distance to the picture height at which the picture quality of each system always falls into the "excellent" evaluation range. The relative viewing distance to the picture height is an alternative expression of FOV for the same television system. These expressions are related as follows.

$$\tan(a/2) = r/(2n)$$

Here, a is FOV, r is aspect ratio, and n is design viewing distance to picture height.

Recommendation ITU-R BT.1845 "Guidelines on metrics to be used when tailoring television programmes to broadcasting applications at various image quality levels and sizings" defines the optimal viewing distance as the distance at which the pixel count per visual angle of one minute is one. It lists the optimal viewing distances to the picture height and the optimal field of view for image systems with various pixel counts (Table1).

These two Recommendations suggest that the picture quality of an image system having viewing conditions in which the angular resolution is one pixel per one arc-minute falls into the "excellent" range.

With above idea in mind, the issue of pixel count should be addressed from the following viewpoints.

−   What should the (minimum and maximum) FOV of UHDTV be and for what reason?

−   Should the criterion for determining the picture quality of UHDTV by the pixel count be the same as that in Recommendations ITU-R BT.1127 and ITU-R BT.1845?

TABLE 1

Optimal horizontal viewing angle and optimal viewing distance in image heights (H) for various digital image systems

Image system	Reference	Aspect ratio	Pixel aspect ratio	Optimal horiz. viewing angle	Optimal viewing distance 1)
720 × 483	Rec. ITU-R BT.601	4:3	0.88	11°	7 H
640 × 480	VGA	4:3	1	11°	7 H
720 × 576	Rec. ITU-R BT.601	4:3	1.07	13°	6 H
1 024 × 768	XGA	4:3	1	17°	4.4 H
1 280 × 720	Rec. ITU-R BT.1543	16:9	1	21°	4.8 H
1 400 × 1 050	SXGA+	4:3	1	23°	3.1 H
1 920 × 1 080	Rec. ITU-R BT.709	16:9	1	32°	3.1 H
3 840 × 2 160	Rec. ITU-R BT.1769	16:9	1	58°	1.5 H
7 680 × 4 320	Rec. ITU-R BT.1769	16:9	1	96°	0.75 H

2.5.3 Viewing distance

Absolute viewing distance differs according to the kind of application. The range may be wider than that of current television because of the variety of applications. Relative distance varies with the field of view and resolution. It also determines the display size and vice versa.

2.5.4 Surrounding luminance

Surrounding luminance affects human perception of displayed images. Therefore, certain assumptions may be needed to determine the system parameters. These assumptions would depend on the kind of application.

2.5.5 Black and white luminance and electro-optical transfer function (EOTF)

The perceptual limit of the black level depends on the accommodation situation of the human visual system (HVS) and is mainly affected by the surrounding luminance. Modern (non-CRT) displays have various native EOTFs, and they can easily be converted into almost arbitrary values by using modern digital techniques. Therefore, the major criterion for determining EOTF may be the efficiency of digital code usage.

2.5.6 Technical constraints imposed by the assumed display device

In the CRT era, system designers tacitly agreed upon the temporal sampling aperture, chromaticity coordinates of tri-stimulus (also their spectrum distribution), and EOTF. The UHDTV display will likely be some other kind of device; hence, those assumptions should be reviewed.

2.6 Feasibility of technologies required for development of UHDTV

In [Input, 9] a broadcaster gave their view on the feasibility of technologies required for UHDTV delivery as follows.

–   It will probably never be possible to implement a UHDTV broadcasting service in the frequency bands currently assigned to terrestrial broadcasting, since it will not be possible to achieve the amount of bit-rate reduction necessary to fit the data rate required by UHDTV into the capacity of current terrestrial broadcasting channels.

–   It may be possible to implement a 3 840 × 2 160 UHDTV broadcasting service in the 12 GHz satellite-broadcasting band in the medium future, when more efficient source-coding and modulation methods will have been developed and implemented in a reliable and viable way. Some experts think that this goal may be attained in about ten years' time, based on studies currently under way on improved source coding and modulation methods. It remains to be seen whether, at that time, the 12 GHz band will be too densely populated to accommodate the new UHDTV services.

–   It will probably be possible to implement a 7 680 × 4 320 UHDTV broadcasting service in the 22 GHz broadcasting service in the far future, when another quantum leap in the efficiency of source coding and modulation methods will have been developed and implemented. Some experts think that this goal may be attained in perhaps 10 to 20 years. It remains to be seen how the 22 GHz satellite broadcasting band will be structured at that time, and which applications will be earmarked for it.

A performance evaluation of compression coding that may be used in the secondary distribution [Input, 17] shows that

–   For 75% of the sequences chosen: DSCQS ≤ 12%

–   For the rest: DSCQS ≤ 30%.

In [Input, 39] the increase in bit rate needed for the picture quality improvement including motion portrayal and colour/tone rendition is discussed from the feasibility point of view as follows.

Since 3 840 × 2 160 UHDTV broadcasting will require a new television receiver, it seems logical to also ask whether the opportunity should also be seized to improve some other aspects of the television image.

Beyond the improvement in resolution, at least two other possible areas of picture improvement could be considered: improvement in colour rendition and improvement in movement rendition.

Since any improvement in service quality has an associated cost, the issue is to decide whether the improvements are worth its cost.

In the case of image resolution, the transition from HDTV to 3 840 × 2 160 UHDTV broadcasting would imply the cost of carrying four times more pixels to the user, which would result in a four-fold increase in uncompressed bit rate. Today, each HDTV broadcast service uses all or a major part of a broadcast emission channel. A four-fold increase in uncompressed bit rate would thus be well beyond the capabilities of the source coding and channel coding methods currently used for television broadcasting. In other words, the emission of a 3 840 × 2 160 UHDTV broadcasting service would require a bit rate that exceeds the capacity of a current terrestrial television channel

In the case of movement rendition, the first possibility would be to double the picture rate by delivering progressive signals rather than interlaced signals. The second possibility would be to further double the picture rate by delivering progressive signals at a picture rate of 100 Hz or 120 Hz, rather than 50 Hz or 60 Hz. Taking into account that the correlation among spatially co-located pixels increases as their temporal distance decreases, each one of those increases in pixel count would probably result in about 50% increase in compressed bit rate assuming no technology advances.

In the case of colour rendition, Working Party 6C has extensively discussed this topic, concluding that an improvement in colour rendition to encompass the complete gamut of visible colours would require at least 10 bits/sample, rather than the 8 bits/sample now required for SDTV and HDTV emission: a 25% increase in bit rate.

It is clear that the possible image improvements above, as well as any increase in the number of audio channels, will all compete for the available emission bit rate, with the four-fold increase in pixel count of 3 840 × 2 160 UHDTV. It will be necessary to make a choice among the improvements that can be fitted in the available emission bit rate, and those that cannot be, unless an improvement is made in source coding, and/or in channel coding, or unless multiple terrestrial TV channels are devoted to a single UHDTV program stream.

3   Baseband image format

3.1   General considerations on the baseband image format in programme signal chain

The baseband image format is the basis of every part of the programme signal chain. Therefore, it should be established first so that every part of the UHDTV broadcast system can be based on it. The formats for different parts of the programme chain may differ from each other because of the requirements for each part. In particular, the acquisition and production format may need some quality headroom for post-processing management.

Recommendation ITU-R BT.1662 "General reference chain and management of post-processing headroom for programme essence in large screen digital imagery applications" describes the general reference chain from acquisition to presentation and recommends as follows (see also Fig. 4, taken from Recommendation ITU-R BT.1662):

−   a single native format depending upon the quality level should be used for programme essence throughout the chain, if possible;

−   if that is not possible due to complex post-processing requirements, then a minimum number of native formats should be used along the chain, and their parameter values should stand in a simple relationship to each other.

The acquisition block should normally use the same native format that is planned to be used for the finished master (simple post-production). However, depending on the nature of required post-production processes, the required amount of headroom on the required parameters should be provided at the input of the post-production block (complex post-production). This can be done by using different native signal formats. To fully benefit from this, the native formats should be related in a direct way to the format intended for the finished master.

This signal chain would likely be applicable to UHDTV since it will deal with programmes having high picture quality. This observation leads to the following suggestions.

−   A baseband image format for high-picture-quality programme production should be developed.

−   It should be based on the image format for finished masters.

FIGURE 4

**Reference chain for typical LSDI applications and cognitive evolution of
picture/sound post-processing headroom along the chain**
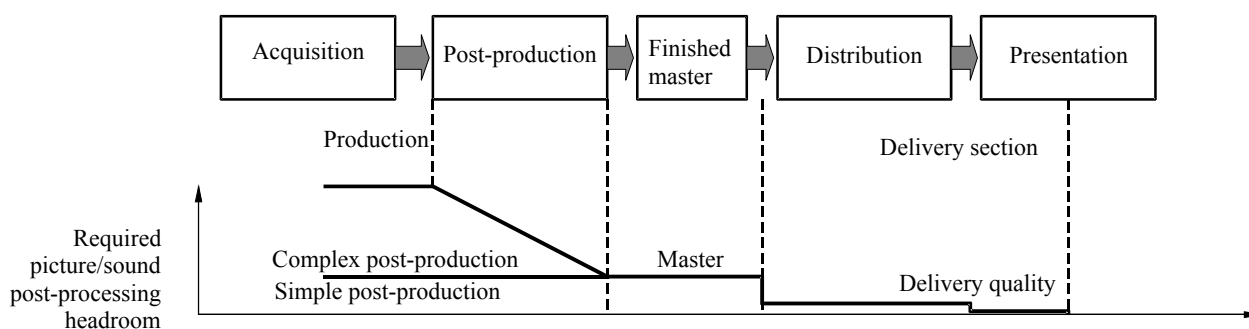
3.2   Parameters

A set of parameter values for the baseband image format was proposed in 2009 [Input, 18], this proposal was further enhanced during the May 2011 series of meetings when proposals from two proponents were merged into a single proposal.

3.2.1   Aspect ratio

Aspect ratio is one of the fundamental properties of the picture because it is difficult to convert one ratio into another without perceptible distortion. A paper in 1990s [1] concludes based on a thorough study that there is no clear evidence of an aesthetic or physiological reason to choose any one aspect ratio over another.

The resolution of EHRI is described in Recommendation ITU-R BT.1201 as a simple integer multiple of 1920 × 1080. This leads to the choice of an aspect ratio of 16:9.

3.2.2   Spatial and temporal sampling structure

A non-orthogonal structure would fit the HVS characteristics. Such a sample structure would result in a reduction in the data rate. Still, continuing the practice of using orthogonal sampling may be desirable. For temporal sampling, interlace would not be needed to reduce the bandwidth because alternative techniques such as high-efficiency video coding are now available.

3.2.3   Pixel count

The pixel count can be derived from the required field of view and required resolution (pixels per unit visual angle).

3.2.4   Frame frequency

3.2.4.1   General

Motion reproduction is one of the most important features of television. If we take "artefact" to mean any perceptual difference between the real scene and the reproduced image, the major artefacts would be flicker, motion blur, and jerkiness (stroboscopic effect).

Flicker is a phenomenon in which the whole screen or a large area of the screen "flickers" due to low frame frequency. Motion blur is due to the accumulation mechanism of the capturing devices. It also happens when moving images are tracked on the screen of a hold-type display. Non-smooth reproduced motion at low frame frequencies is called jerkiness. The use of the shutter during acquisition can produce a multi-exposure-like image even at high frame frequencies. This phenomenon is sometimes called jerkiness as well. Motion blur and jerkiness depend on the object speed.

The object speed (arc-degree/second) generally tends to become faster as the field of view of the system becomes wider. For example, given the same shooting angle, an object on an HDTV screen will move faster than the one on an SDTV screen. This implies that high definition TV systems have stricter motion portrayal requirements.

3.2.4.2   Studies done in the HDTV R&D era

The study on the frame frequency was the most difficult one to reach an agreement on in the HDTV standardization. Report ITU-R BT.801, The present state of high-definition television, describes it in detail.

The study was concerned with two major factors.

−   motion portrayal;
−   the relationship with film and with current and future TV systems.

After quite a lot of discussion, consequently, 60, 50, 30, 25, and 24 Hz were adopted as possible frame frequencies. The first two values are the same as in SDTV whereas the lower three frequencies are intended for special program production.

The rationale for maintaining the same frame (field) frequencies as of SDTV regardless of the desire for motion portrayal improvement was as follows.

−   Dynamic resolution is improved by the use of a shutter in the camera.

Regarding the desire for flicker reduction, the rationale for maintaining the same frame (field) frequencies as in SDTV was as follows.

−   It is not essential that the display refresh rate be the same as the studio or the emission signal field-rates.

The current HDTV has a 50 or 60 Hz display refresh rate, 100% aperture in acquisition, and virtually 0% (impulse-type) aperture in display.

The recent advent of FPDs that have a long hold time have solved one problem at the expense of causing another. Their hold mechanism works favourably to reduce flicker. In fact, large-screen FPDs have no flicker problems. However, the hold mechanism also creates motion blur. Many studies have been undertaken to seek a means to eliminate motion blur. So far, having the display itself create and insert frames to double, triple, or even quadruple the frame frequency is seen as a potential solution. It is also noted that newer FPD technology does not suffer from long hold times

Our review of the studies on the frame (field) rate of HDTV suggests that we should look carefully at the UHDTV frame rate from the following aspects.

−   Widening of the field of view (FOV) may affect the human visual system's perception of the portrayed motion. The critical flicker frequency increases as FOV becomes wider, for example.

−   How much improvement is worthwhile?

−   The results of such investigations depend on the type of display and signal processing at the display itself.

3.2.4.3   Recent studies

Flicker [2]

Critical flicker frequencies (CFFs) with varying duty ratios and FOVs were measured in order to characterize the parameters of the HVS while viewing a wide FOV display. Two experiments were conducted. In one experiment, CFFs were measured for various duty ratios ranging from 30% to 90% with the FOV of 100 arc-degrees. In the other experiment, CFFs were measured for 30 and 100 arc-degree FOVs under the condition of the most sensitive duty ratio for each viewer. These two experiments enabled the CFFs of wide FOV systems to be compared with those of narrow FOV systems.

Figure 5 shows the mean CFFs and standard deviations at FOV of 100 arc-degrees for duty ratios of from 30 % to 90 % at three levels of luminance. CFFs decreased from 80 Hz to 65 Hz as the duty ratio increased.
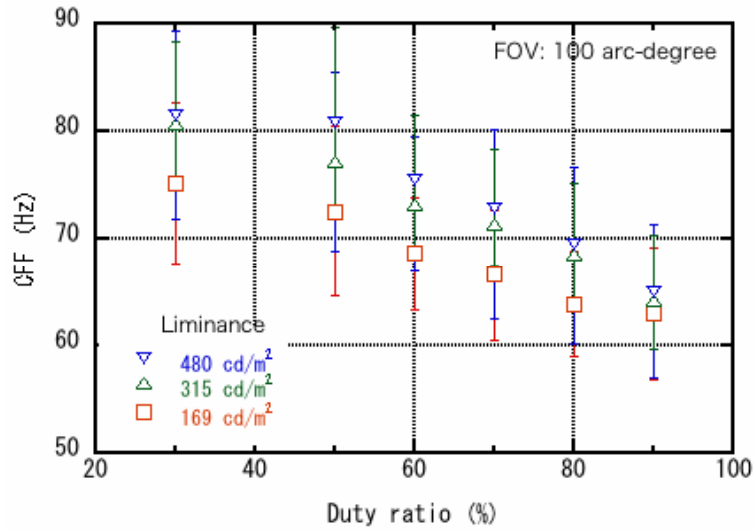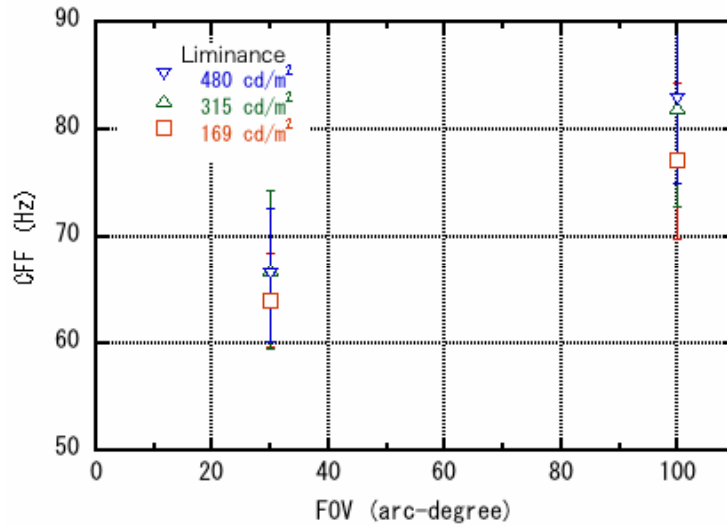
FIGURE 5

Relationship between CFF and duty ratio

Figure 6 shows the mean CFFs and standard deviations for FOVsof30 and 100 arc-degrees at three levels of luminance. The CFFs of the wide FOV were above 80 Hz and were higher than the CFFs of the narrow FOV (about 65 Hz).

FIGURE 6

Relationship between CFF and FOV

These results show that CFFs decrease as the duty ratio increases. This suggests that viewers would perceive more flicker in impulse-type displays than in hold-type displays, and that flicker perception (or the decrease in image quality caused by flicker) might be a problem specific to wide-FOV display systems. On the other hand, as mentioned above, hold-type displays have perceptible motion blur. We should optimize the trade-off between perceptions of flicker and motion blur at a certain frame frequency. Furthermore, we believe the best way to resolve this trade-off is to increase the system frame frequency. It will thus be essential to determine an adequate system frame frequency for wide FOV systems.

Motion blur and stroboscopic effect [3]

Motion blurs at various aperture times and object speeds were evaluated. One of the scenes that give viewers a very unnatural feeling is when the perceived sharpness of an object changes significantly between conditions of stillness and motion. To emulate such a scene, the test involved sliding a still picture into the field of view, stopping it at the centre for three seconds, and then sliding it out (Fig. 7). The test materials are shown in Fig. 8. The sliding speed was varied up to 32 arc-degrees/second to match the characteristics of camera panning in actual programs. The moving part of the test materials was subject to a moving average filter so that accumulation at the camera aperture of the imaging device or at the retina in the display aperture would be simulated. The viewers evaluated the sharpness of the moving part in comparison with the still part as a reference on a 5-grade impairment scale.

The results are shown in Fig. 9. The horizontal axis represents the number of the pixels of the filter that operated on the moving part. The three lines are for object speeds of 32, 16, and 8 arc-degrees/second, respectively. The evaluation for long filters becomes less severe as the object speed increases. 2~3 pixels/frame gives the perceptible limit, and 6~11 pixels/sec gives the acceptable limit in Fig. 9a. 1~2 pixels/frame gives the perceptible limit, and 6~11 pixels/sec gives the acceptable limit in Fig. 9b.

FIGURE 7

Sequence of evaluation stimuli
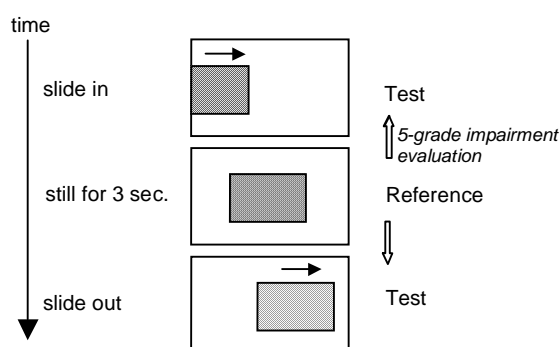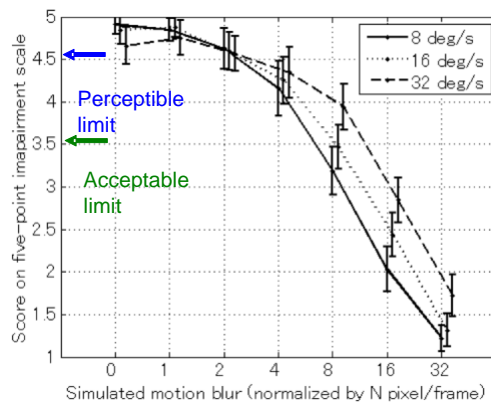
FIGURE 8

Test materials for blur evaluation

Woman   Yacht harbor

FIGURE 9A

Results of subjective evaluation of motion blur

Woman

FIGURE 9B

Results of subjective evaluation of motion blur
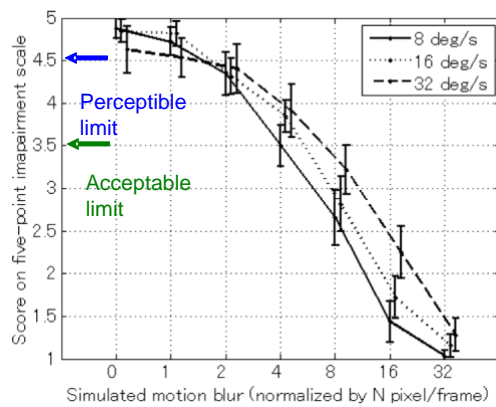
Yacht harbor

Figure 9c shows the relationship between time aperture and object speed giving an acceptable level of motion blur derived for the data in Fig. 9a and b [6C/456].

FIGURE 9C

Relationship between time aperture and object speed giving an acceptable level of motion blur
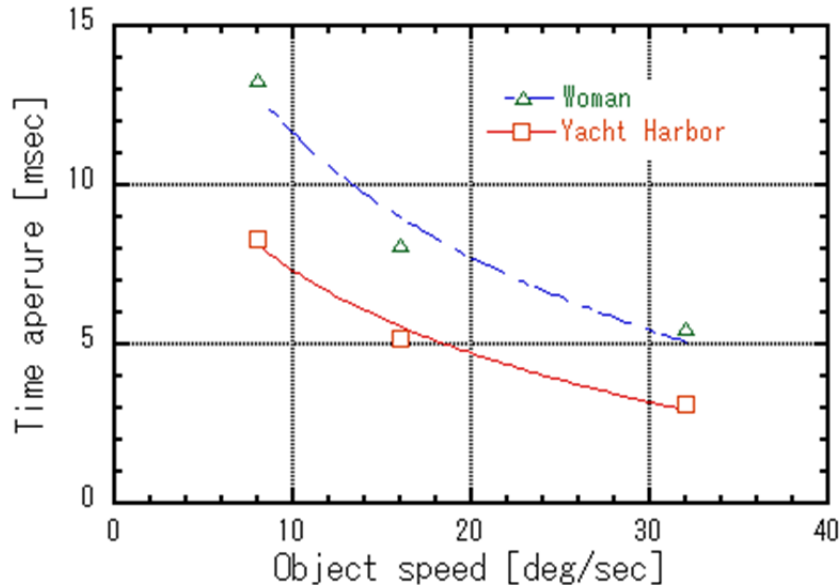
Stroboscopic effects at various frame frequencies and aperture times were evaluated. It is known that a camera shutter causes a stroboscopic effect. The relationship between the effect and temporal sampling parameters has been studied. This stroboscopic effect is sometimes called jerkiness. It may be referred to as "multiple images" that appear when the frame frequency is high relative to the reciprocal of the "time constant" of vision. In other words, the stroboscopic effect arises when the image of the object that is not or cannot be tracked by the eye moves discretely across the retina. The stroboscopic effect frequently appears in sports programs. Therefore, the test materials were chosen from such a genre and were taken by a high-speed HDTV camera with various temporal parameter values (Fig. 10). The object speed was varied from 20 to 170 arc-degrees/second.

Figure 11A shows the results for sequences with a constant aperture time of 1/240 second and different frame frequencies. There was a degradation of 1 rank at 120 Hz/50%, 1.5 ~ 2 ranks at 80 Hz/33%, and 2 ~ 2.5 ranks at 60 Hz/25% compared with the reference at 240 Hz/100%. Figure 11b shows the relationship between image quality score and frame frequency for stroboscopic effect derived for the data in Fig. 10 (See doc. 6C/456).

Figure 12 shows the results for a constant frame frequency of 240 Hz and different aperture times. It shows that the stroboscopic effect appears even at a frame frequency of 240 Hz for high-speed objects, whereas it is not apparent for low-speed objects.

FIGURE 10

Test materials for stroboscopic effect evaluation[1]

Runner
20 deg/sec

Stadium
25 deg/sec

Tennis
180 deg/sec

Pitcher
140 deg/sec

---

[1] They were captured using an HDTV high-speed camera with a frame frequency of up to 240 Hz and various shutter speeds. The maximum object speed is noted.

FIGURE 11A

**Results of subjective evaluation of stroboscopic effect,
constant aperture time: 1/240 sec.**
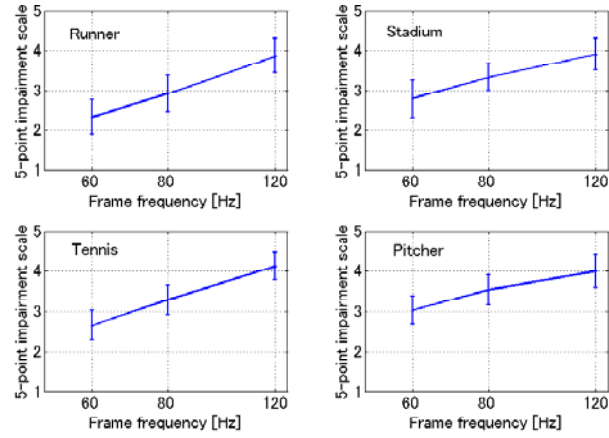
FIGURE 11B

Relationship between image quality score and frame frequency for stroboscopic effect
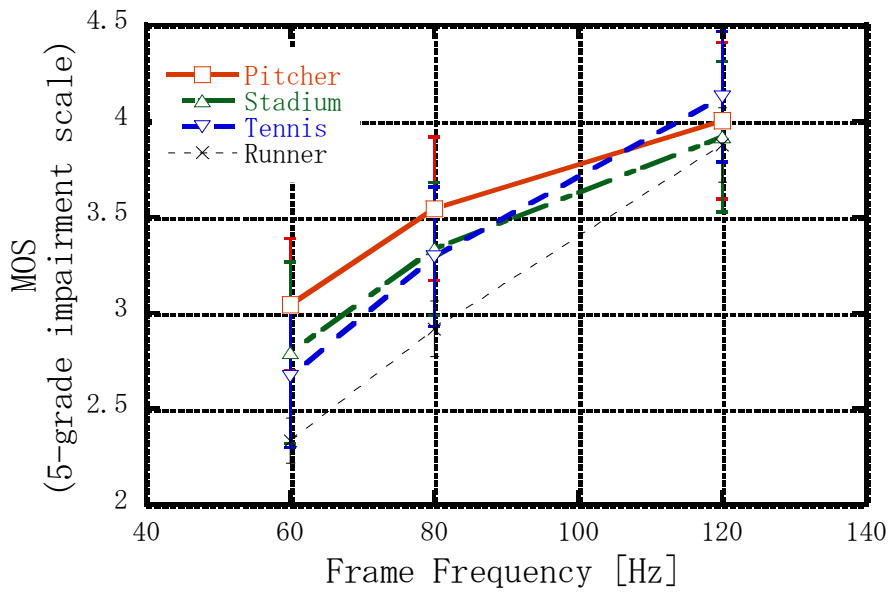
FIGURE 12

**Results of subjective evaluation of stroboscopic effect,
constant frame frequency: 240 Hz**
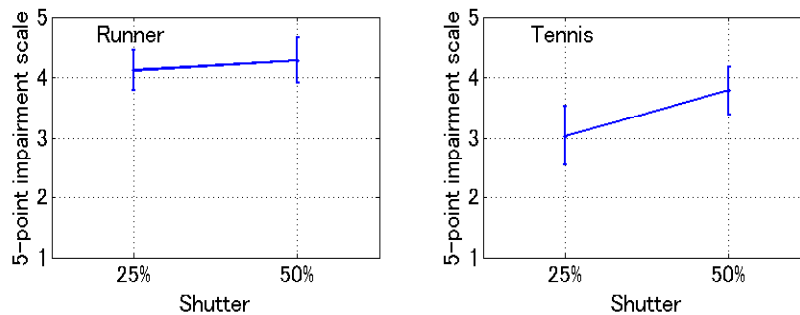
Moving pixels per aperture time can be used as an index for motion blur, and moving pixels per frame time can be used as an index for the stroboscopic effect. The indices of 1.2 pixels/frame and 6 pixels/frame give the perceptible limit and acceptable limit for motion blur, respectively. They correspond to the accumulation times of 1/1600 s and 1/320 s, respectively. These values seem slightly more demanding than the previously reported results do. The reason can be found in the comparison procedure used. Our procedure may make it easier than the previous ones to perceive differences. Our results suggest that the stroboscopic effect might not be perceived when the object motion is below 5 to 10 pixels/frame. The difference in the perceptible limits between motion blur and the stroboscopic effect may be attributed to the difference between fovea and peripheral viewing.

The index of pixel/frame is determined from the frame frequency, accumulation time, and object speed. Therefore, how a fast-moving object is dealt with is an important consideration in the system design. Data obtained from analyses of broadcasting programs would be useful for studies aimed at improvement of current systems. On the other hand, a study aimed at development of a new system should make clear what kinds of objects are dealt with and how are they expressed on the screen.

In addition, it should be taken into account that another picture parameter affects the human visual system, and hence the requirement for motion portrayal changes. It has been pointed out that dynamic visual acuity (DVA) is affected by the field of view (FOV) through which the moving target is presented [4]. Evidence for this effect is shown in Fig. 13. The vertical axis in the figure is the object speed above which viewers could not recognize the object's orientation. DVA, in this case the viewer's ability to resolve an orientation of a moving target (Landolt C), increases by a factor of 1.8 as the FOV increases from 10 to 110 arc degrees.

FIGURE 13

Measurements of DVA
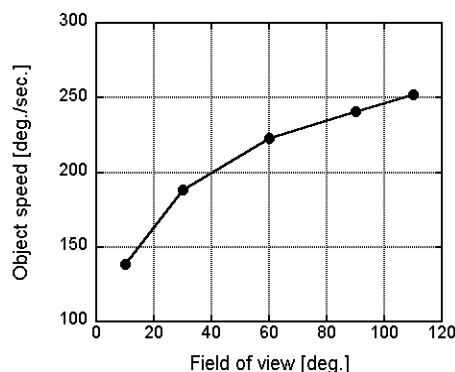
The motion portrayal of television depends on its scanning scheme and temporal sampling parameters. The parameters comprise the frame frequency and aperture time. Aperture time has not been explicitly included in television signal standards, but general program productions have been done with an aperture ratio of 100% at a camera and virtually 0% at a display because of technical constraints. The aperture ratio is thus one of the most important aspects in envisaging the limits of motion portrayal and determining the temporal sampling parameters of new high-quality television systems beyond the level of HDTV. A CRT will not be the only display device for such a future system. Inevitably, a hold-type display should be considered.

The above subjective tests were conducted with test materials similar to those used in daily broadcasting with the aim of establishing guiding principles for the design of future high-quality television systems. The results presented here can be taken as guiding principles when designing new television image formats and developing cameras and displays.

3.2.4.4 Proposed higher frame frequency

The recent UHDTV image format proposal suggests a frame frequency of 120 Hz as a higher frame frequency for UHDTV for the following reasons: (See doc. 6C/456)

−   The frame frequency of 120 Hz eliminates the flicker artefact completely. This means that the shorter the light emitting time of displays during the frame period, the more the motion blur caused by the hold-mechanism is reduced without introducing flicker. The frame frequency at the camera end keeps the stroboscopic effect within the acceptable limit. The use of an electronic shutter of approximately a half to one third keeps motion blur within the acceptable limit. These results show that the frame frequency of 120 Hz at least is required for a worthwhile improvement in motion image quality, although higher frame frequencies than 120 Hz can yield further improvements for more demanding criteria.

−   Some UHDTV applications may require the same frame frequencies as those of HDTV, which are up to 60 Hz. In defining frame frequencies higher than 60 Hz for UHDTV, it would be reasonable to propose a unique value. From purely a technical perspective it is desirable to adopt the least common multiple (LCM) frequency of the existing television systems frame frequencies for easy conversion. But the value of 300 or 600 Hz does not seem reasonable from the standpoint of balancing achievable performance with technical difficulty and the required bandwidth in the foreseeable future. Consequently, a conversion from/to 120 Hz to/from 50 Hz or 59.94 Hz would have to be made. These

frame conversion problems could be solved by using modern sophisticated digital technologies.

– The feasibility of the camera, display, and transmission technologies needs to be considered before setting a higher frame frequency. Recent progress would indicate that a UHDTV system with a frame frequency of 120 Hz will be feasible sometime in the near future.

3.2.4.5 Subjective evaluation of moving picture quality [Input, 47]

A recent paper has reported the subjective evaluation results of moving picture quality on a large screen [2]. An HDTV high-speed camera and projector that operate up to 240 Hz were used. Common images for television programs were used as the test materials. The set of test materials included sports scenes, such as football, baseball, tennis, and running, and several scenes taken at an amusement park. Test conditions are listed in Table 2. Adjectival categorical judgement with a five-grade quality scale was used. Figure 14 shows the evaluation scores converted to an interval scale for each test material. The results show that the scores increase as the frame frequency becomes higher while how much the picture quality improves differs from material to material. The range of score improvement between 60 and 120 Hz extends from 0.14 to 1.04. The average scores are shown in Fig. 15. The improvement in scores between 60 and 120 Hz, and 120 and 240 Hz are 0.46 and 0.23 respectively. Both differences are statistically significant. It is anticipated that the increase of frame frequency will be more effective for UHDTV that has much more resolution than HDTV. These results lead to the conclusion that a frame frequency of at least 120 Hz is desirable for UHDTV to improve its moving picture quality.

FIGURE 14

Results of subjective evaluations of moving picture quality at high frame frequencies
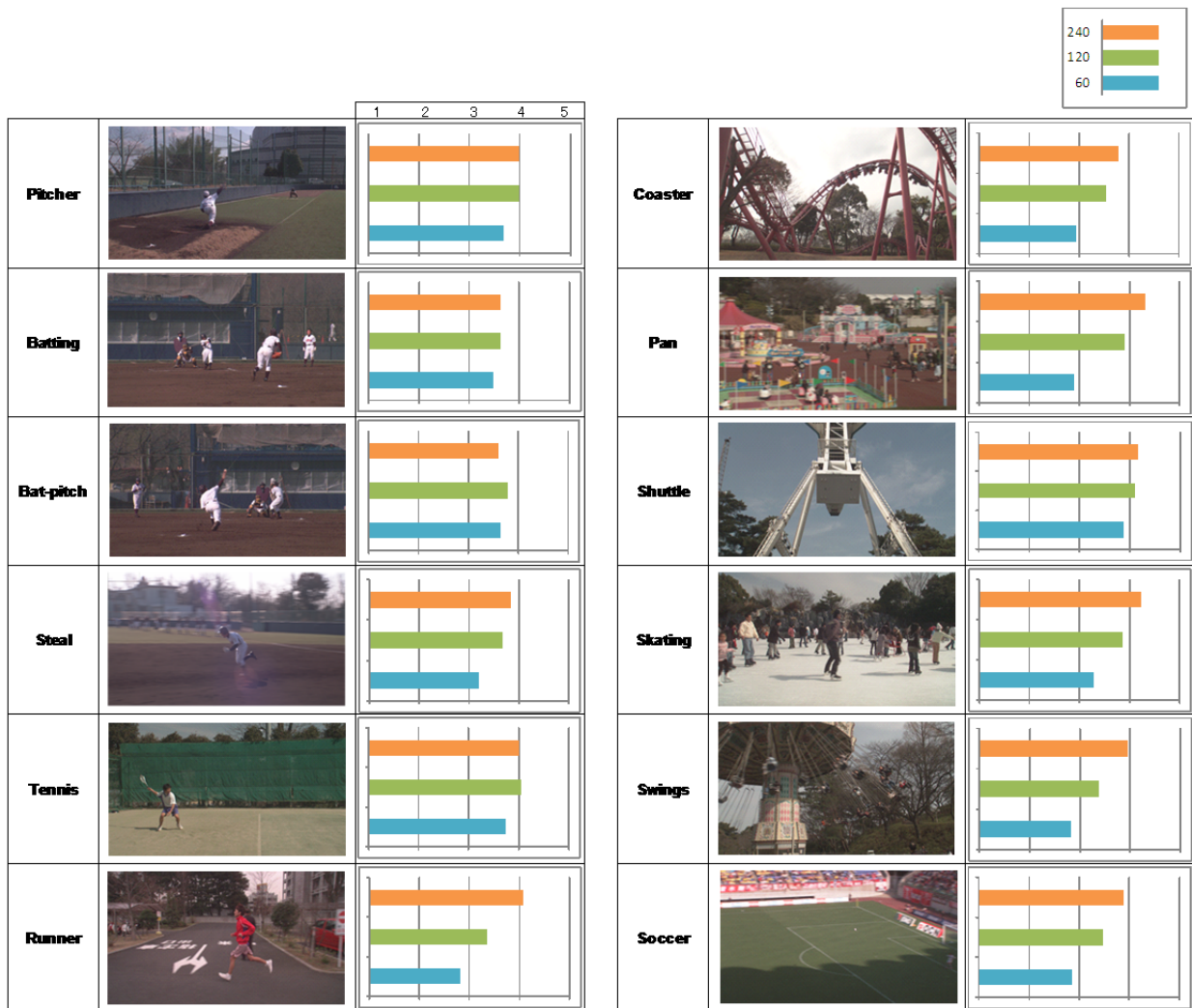
TABLE 2

Test conditions

Resolution	1 920 × 1 080
Scanning	60, 120, 240 Hz, progressive
Screen size	100 inches
Viewing condition	Recommendations ITU-R BT.500, ITU-R BT.710
Viewing distance	3.7 m
Participants	69

FIGURE 15

Average score
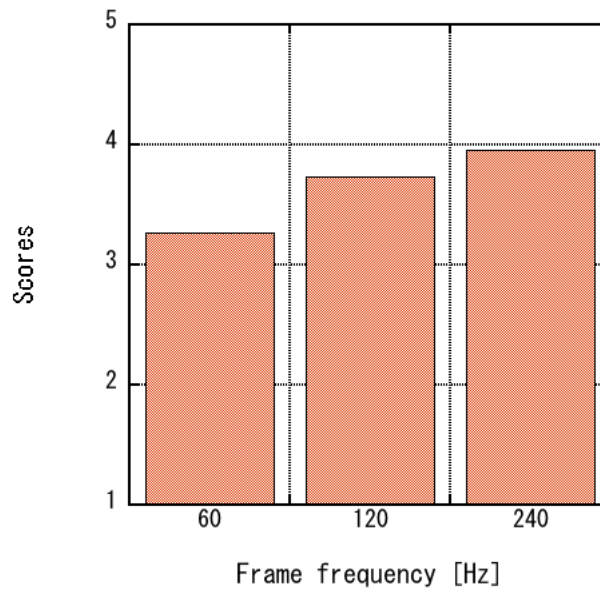
3.2.5   Non-linear coding function

The non-linear coding function, or gamma pre-correction, of video signals has two roles. One is to raise the code-usage efficiency by making the perceptual quantization error among the video codes more uniform (see chapter 3 for a description). This is similar to the one in analogue video to reduce the perceptual effects of noise on the transmission path.

The other role is to pre-compensate for the non-linear electro-optical transfer function (EOTF) of displays. The combination of the gamma pre-correction and the gamma characteristics of the display device yields the total transfer function, or end-to-end gamma, from the scene being shot to the reproduced image on the screen. It is well known that the end-to-end gamma is not necessarily unity; i.e., the gamma value is not necessarily 1. For example, some people believe that values of 1, 1.25, and 1.5 are appropriate for bright, dim, and dark surrounding environments, respectively [6]. Furthermore, the end-to-end characteristics are more complex than a simple power function because the optical-electro transfer function (OETF) at the acquisition end consists of a power-function segment and a linear segment. The current tone curve is analysed below with this point in mind.

[Input, 52], Proposal of required electro-optical transfer function for a FPD as a master monitor, describes the measured EOTFs of current HDTV reference monitors. According to that document, the EOTF of a CRT can be represented by Equation 1.

$$L = k\,[V + b]^{\alpha} + \beta \tag{1}$$

where,

L: Screen luminance (normalized)

V: Input video signal level (normalized)

α: Inherent gamma of the device

β: Luminance with no signal input

b :   Variable for brightness control

k :   Coefficient for normalization

$[x] = x$ when $x >= 0$, $[x] = 0$ when $x < 0$

The variable "b" represents the brightness control, and it is adjusted to conform to the viewing environment by using the procedure described in Rec. ITU-R BT.814. Figure 16 plots five EOTFs with different brightness control settings.

FIGURE 16

Electro-optical transfer function model
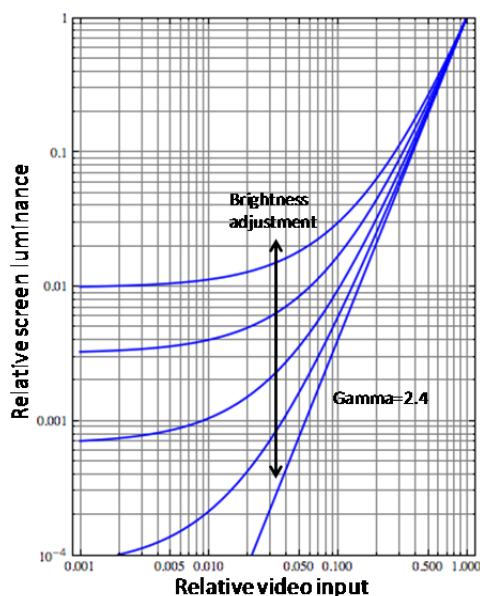
The gamma pre-correction described in Rec. ITU-R BT.709 is applied to HDTV signals (Fig. 15). Modifications for artistic reasons will not be considered here. Calculated end-to-end characteristics are described in Fig. 18. Figure 18 illustrates that the brightness adjustment according to the viewing environment makes the end-to-end gamma vary so as to match with the visual perception predicted by the colour appearance model. Namely, increasing the brightness in bright-surroundings makes the end-to-end gamma close to unity, whereas decreasing the brightness in dark surroundings makes it greater than unity. The end-to-end gamma is an intermediate between these two values in dimly lit surroundings.

FIGURE 17

Optical-electro transfer function of Rec. BT.709
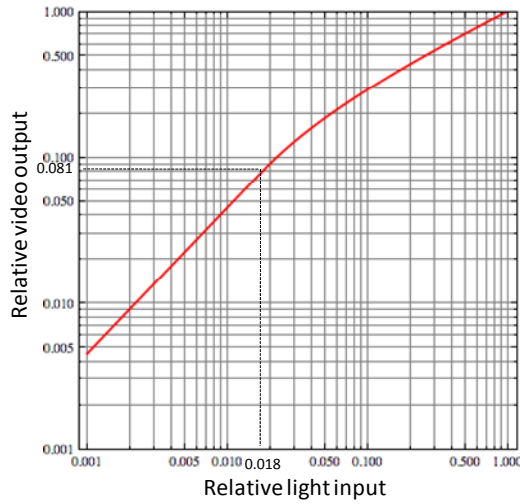
FIGURE 18

End-to-end transfer function: OETF with linear segment
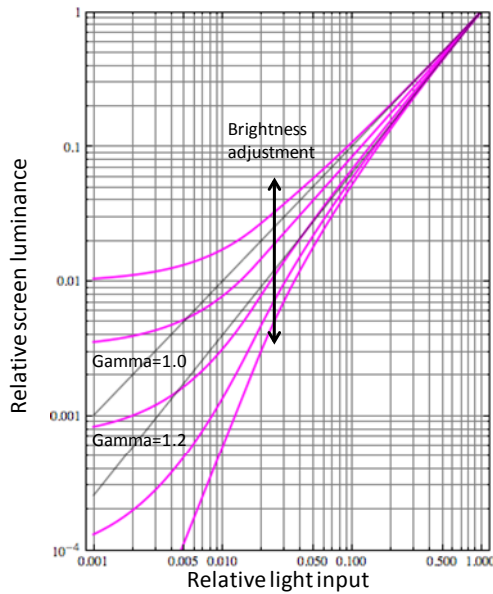
Figure 19 plots cases in which the OETF is a simple power function without a linear segment. A comparison between Figs. 18 and 19 shows that the linear segment of the Rec. ITU-R BT.709 OETF has a significant effect at low luminance.

The above discussion has illustrated an ingenious mechanism of the combination of OETF of Rec. ITU-R BT.709 and the EOTF of CRTs. It unintentionally makes the end-to-end tone curve quite optimal.

FIGURE 19

End-to-end transfer function: OETF without linear segment
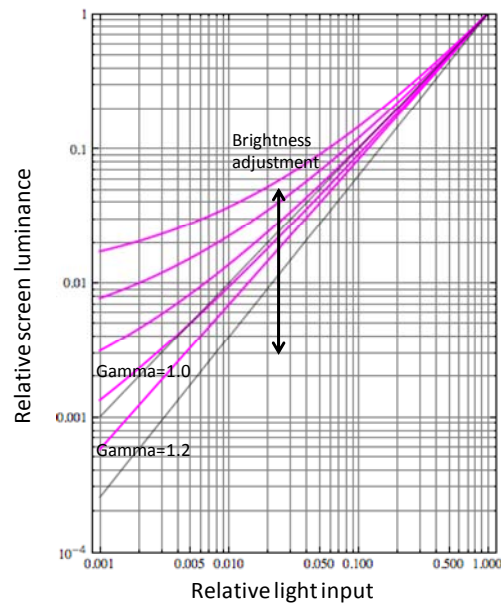
3.2.6 Colorimetry

The first steps to determine the UHDTV colorimetry should address the following questions.

– What are the colour rendition requirements?

– Are there other colorimetry requirements?

– What method and parameter values should be chosen in light of those requirements?

One of the points to be discussed is whether the colorimetry should be based on the display tri-colours or not. Expansion of the colour gamut will enable UHDTV to provide a better visual experience to viewers. Table 4 classifies the methods to expand the colour gamut.

One proposal [Input, 10] is that UHDTV should cover the entire real surface colours with good efficiency.

The chromaticity coordinates of primaries and reference white are proposed in [Input, 16] and [Input, 18]. They are listed in Table 3.

TABLE 3

Proposed sets of primaries and reference white

Parameters		Values		
		x	**y**	**Wavelength (nm)**
Chromaticity coordinates (CIE 1931 xy) of reference primaries and reference white	Redprimary	0.7006(a)	0.2993(a)	625
		0.7140(b)	0.2859(b)	635
	Greenprimary	0.1625(a)	0.8012(a)	531
		0.1702(b)	0.7965(b)	532
	Blueprimary	0.1314(a)(b)	0.0459(a)(b)	467
	Reference white (D65)	0.3127(a)(b)	0.3290(a)(b)	-

Sets (a) and (b) are proposed in [Input, 18] (Korea) and [Input, 16] (Japan), respectively.

Both proposals are based on the following concept.

−   The colour gamut for UHDTV should be expanded by using wider colour primaries than conventional ones so that it covers real surface colours as much as possible.

−   Those primaries should be real colours.

As a result, both proposals select monochromatic colours as RGB primaries. The arguments are summarized as follows, in light of both proposals on the derivation of the primaries.

−   531 nm for green and 625 nm for red are on the equi-hue lines of the current LED-backlit LCD and AM-OLED. However, 635 nm is far located from the red primaries of the current LCD and AM-OLED technology.

−   532 nm for green and 635 nm for red are easy to realize with the current laser technology. On the other hand, at present it is very difficult to manufacture 625 nm and 531 nm LDs.

−   635 nm for red has less energy efficiency because human visual sensitivity is lower at 635 nm than at 625 nm.

3.2.6.1   Examination of two sets of RGB primaries [Input, 24]

Introduction

Two RGB-primary sets were introduced in WD-PDNR ITU-R BT.[Image-UHDTV] [Input, 19]. Table 3 describes their chromaticity coordinates in the CIE xy domain and their corresponding wavelength (nm) on the spectral locus. Figure 20 shows the two RGB-primary sets and their constant hue lines which were computed using the CIELAB hue. The largest difference can be found in the red primary position. The 625 nm position was chosen in order to encompass efficiently most of the real-world surface colours and also by taking into account the characteristics of flat panel displays such as AMOLED and LCD on which UHDTV programs will be presented. (see Appendix 1 in [Input, 19]). The 635 nm position was selected by considering laser display technology (see Appendix 2 in [Input, 19]).

FIGURE 20

**The two sets of RGB primaries introduced in Table 3 and their constant hue lines
in the CIE xy coordinates**
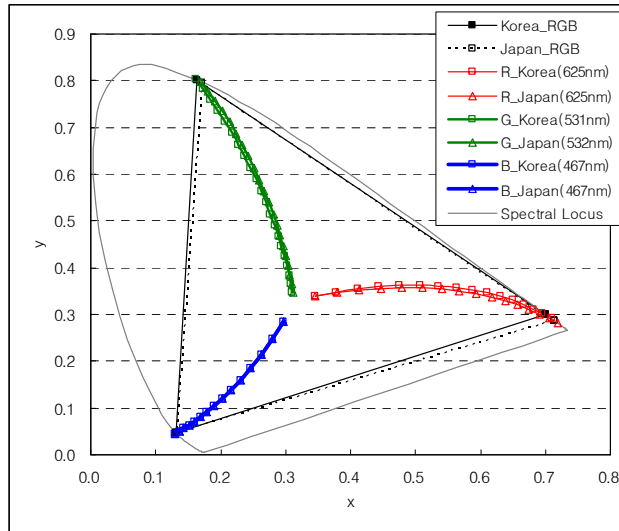
Table 4

Classification of methods to expand the colour gamut of image systems

Approach	Name	Description	Performance	Implementation and compatibility
Display referred approach	Negative RGB	Use negative or more than unity values for RGB while using conventional RGB primaries e.g. Rec. ITU-R BT.1361, xvYCC	Considerably wide, but not all in some case, part of the area within colour locus will be covered Optimal code usage efficiency may be difficult	Easy implementation of relatively small circuitry in mixed environment with current systems Some standard or guideline that describes a real colour gamut expanded display may still be needed* Incompatibility with positive-value-based signal processing
	Wider RGB (real colour)	Use widened RGB primaries comprising real colours e.g. Adobe RGB	Considerably wide, but not all, part of the area within colour locus will be covered Code usage efficiency is expected to be high	No conversion circuitry is required after display compliant with the standard is introduced
	Wider RGB (imaginary colour)	Use widened RGB primaries including imaginary colour(s) e.g. [Input, 10] (Korea)	Flexible choice as to cover the target gamut (e.g. real surface colours) Small efficiency drop due to the coverage of outside part of colour locus	Some conversion circuitry is needed at display end * See above
	XYZ based	Use XYZ colorimetry and converted primaries for transmission if needed e.g. [Input, 53] (USA), SMPTE 428-1-2006	Any signal falling onto colour locus is covered Code usage efficiency is not expected to be high	Some conversion circuitry is needed at display end * See above
Scene referred approach	Multiple (more than three) primaries	Use more than three primaries, spectral distribution reproduction ultimately	Environment-independent reproduction	Completely different system

The influence of changes in red primary

To determine the best red-primary position, three unnoticeable areas around three red constant-hue lines of 625, 630 and 635 nm were compared. Figure 21 illustrates these three unnoticeable areas that can be formed by three pairs of red, blue and green lines corresponding to 625, 630 and 635 nm respectively. The colours located in each of three unnoticeable areas fall into the colour difference range of $\Delta E = \pm 2$ (in CIELAB) against each of three constant hue lines. It is generally known that if the colour difference between two stimuli is less than about 2, the two stimuli might not appear to have different colour. As mentioned in the previous section, 625 nm was chosen by considering the characteristics of flat panel displays such as AMOLED and LCD whereas 635 nm was selected by taking into account laser display technology. The unnoticeable area for the red primary of 630 nm (the area generated by two blue lines in Fig. 21) is seen to overlap largely with the two unnoticeable areas for two red primaries of 625 and 635 nm. This suggests that the red primary of 630 nm can reflect the colour characteristics of both types of display – flat panel displays and laser displays. Thus, the red primary of 630 nm can be the best choice amongst 625, 630 and 635 nm.

It can be presumed based on the above investigation that the extent and efficiency in the inclusion of real surface colours are insignificantly affected by changing red-primary position within 625-635 nm. The four candidates of RGB primaries were therefore selected and were evaluated in terms of gamut coverage and gamut efficiency. Table 5 explains the evaluation results for the four chosen candidates. In the computation of the gamut-coverage and -efficiency, the reference gamut was firstly established from the real surface colours (Pointer and SOCS database) and the reproducible colours by the RGB primaries of AMOLED [7], LCD [7], Digital-Cinema Reference Projector [7], and three standard colour spaces (Recommendation ITU-R BT.709 [9], Adobe RGB [10], and NTSC [11]). Details for computing the gamut-coverage and -efficiency were described in Appendix 1 of WD-PDNR ITU-R BT.[Image-UHDTV] [Input, 19]. The calculated results of gamut coverage/efficiency in Table 5 suggest that there is little difference for all four candidates in the aspect of coverage of the colours falling onto the reference gamut.

FIGURE 21

**Three unnoticeable areas around three red constant-hue lines of 625, 630 and 635 nm
in the CIE xy chromaticity diagram**
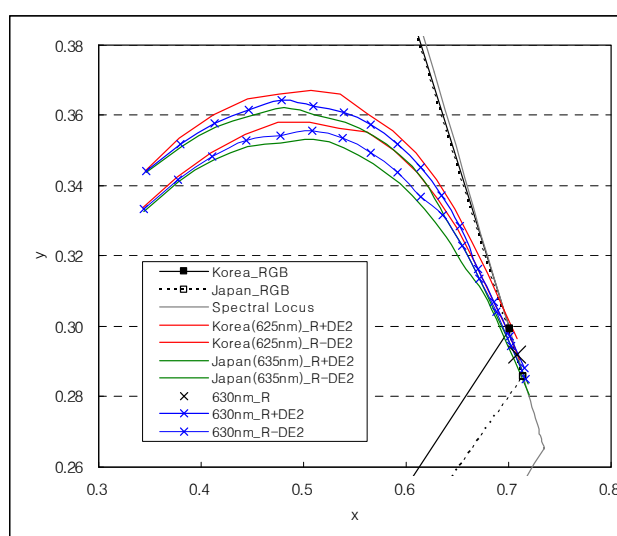
TABLE 5

Four candidates of RGB primaries for UHDTV systems

Primary	Korean Proposal (Appendix 1 in WD-PDNR ITU-R BT.[Image-UHDTV], [Input, 19])	Japanese Proposal (Appendix 2 in WD-PDNR ITU-R BT.[Image-UHDTV], [Input, 19])	Modification 1	Modification 2
Red	0.7006, 0.2993 (625 nm)	0.7140, 0.2859 (635 nm)	0.7079, 0.2920 (630 nm)	
Green	0.1625, 0.8012 (531 nm)	0.1702, 0.7965 (532 nm)	531 nm from Korean proposal	532 nm from Japanese proposal
Blue	0.1314, 0.0459 (467 nm)			
Reference White (D65)	0.3127, 0.3290			
Gamut Coverage	96 %	95 %	95 %	95 %
Gamut Efficiency	91 %	90 %	90 %	90 %

Conclusions

Two RGB-primary sets for UHDTV systems were proposed by Korean and Japanese administrations during the previous ITU-R WP 6C meeting in November, 2009 and were described in WD-PDNR ITU-R BT.[Image-UHDTV] [Input, 19]. A single suggestion was made for blue primary which was placed at 467 nm on the spectral locus. The difference in the two suggested green-primaries was insignificant i.e., 1 nm difference occurring between 531 nm by Korea and 532 nm by Japan. The main difference in the two proposals was the location of red primary: 625 nm by Korea and 635 nm by Japan. The influence of changes in red-primary position was thus examined within 625-635 nm. The red primary of 630 nm was found to be the best choice, because this red primary can reflect colorimetric characteristics of both flat panel displays and laser displays. Two new sets of RGB primaries were then formed and evaluated in terms of gamut-coverage and -efficiency. These were constituted of the same red (630 nm) and blue (467 nm) primaries, but different green primaries (531 and 532 nm). Both two sets performed equally in covering the reference data set that were composed of real surface colours, the reproducible colours of flat panel displays such as AMOLED and LCD and Digital Cinema Reference Projector, and standard colour spaces of Adobe RGB, NTSC and Recommendation ITU-R BT.709. In conclusion, either new RGB-primary set (R 630 nm, G 531 nm, B 467 nm or R 630 nm, G 532 nm, B 467 nm) can be recommended for UHDTV systems.

3.2.6.2 Capture and display experiments of wide colour gamut images [Input, 28]

Objects of shooting

Red and yellow flowers were selected as highly-saturated natural objects commonly used in TV programs. For blue and cyan natural objects, butterflies were selected. A red model car and stained glass tableau containing green as artificial colours are also selected.

Capture

The objects were illuminated by xenon lamps that have the spectral energy distribution close to that of the sunlight in the daytime. The spectral energy distribution was adjusted by spectral correlation filters(SOLAX, SERIC Ltd., Japan).The stained glass tableau was backlit with a fluorescent light.

The tristimulus values XYZ of the objects were measured with a CCD-based imaging colorimeter (PM-1400; Radiant Imaging, USA).

Display

The images data are converted to RGB values on the basis of a laser display (LaserVue, Mitsubishi, Japan) [12] with monochromatic RGB primaries (λB = 447 nm, λG = 532 nm, λR = 640 nm, white point: D65). For comparison of the reproduced colours between UHDTV and HDTV, the RGB images were converted to ones the colour gamut of which were limited to that of the HDTV (Recommendation ITU-R BT.709).

FIGURE 22

Colour distribution of objects on the x-y chromaticity coordinates

(Inner triangle: HDTV primaries, Outer triangle: UHDTV primaries)

(a) Carnation

(b) Geranium and marigold

(c) Sunflower

(d) Butterfly

| (e) Model car | (f) Stained glass |
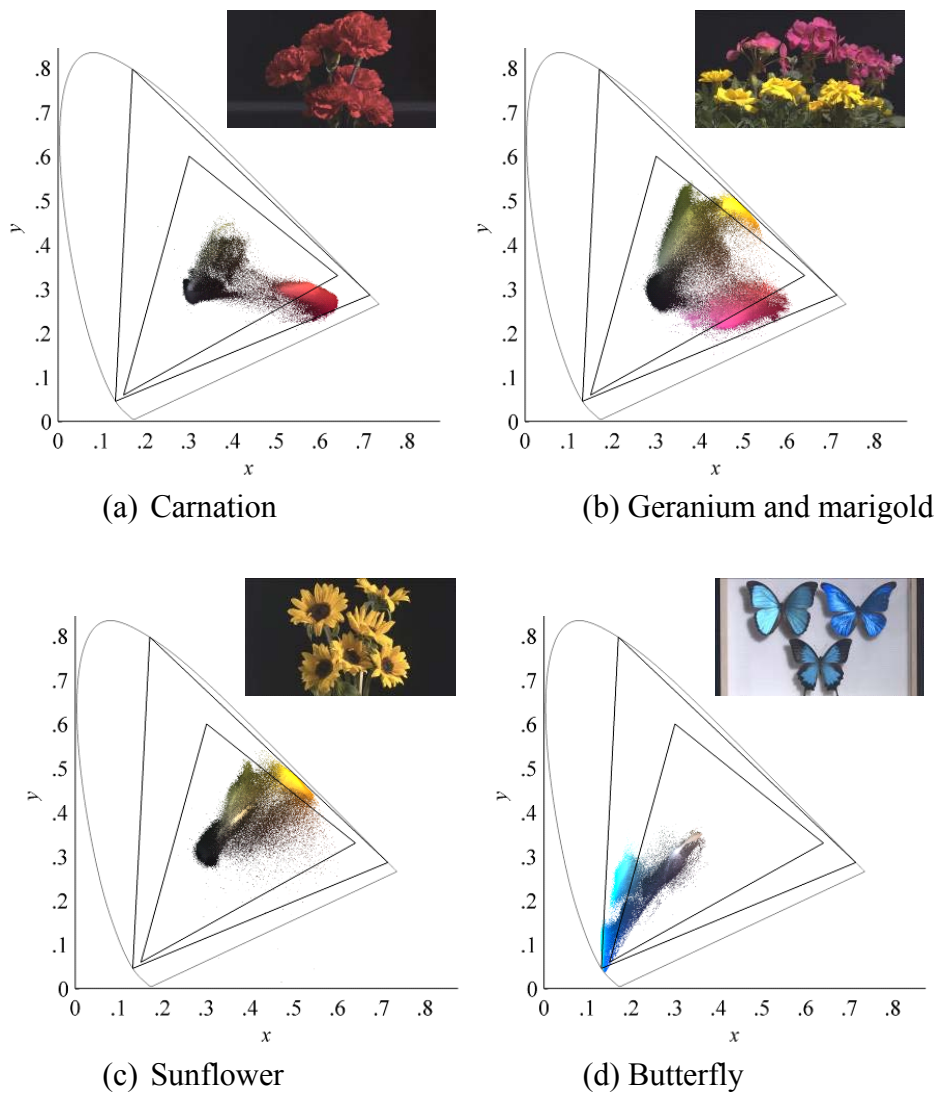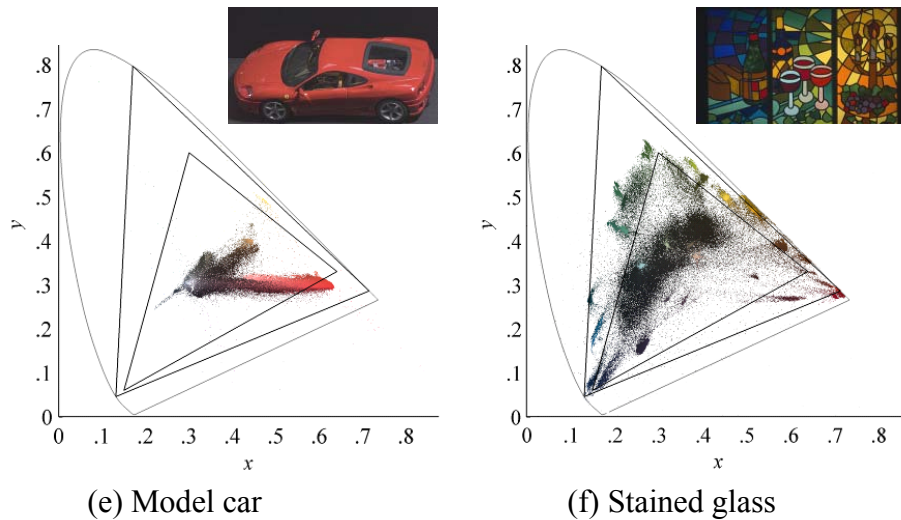
Results

Figure 22 shows the colour distribution of each object on the x-y chromaticity coordinates. It was found that some objects commonly used in TV programs contain highly saturated colours that are not reproducible with the HDTV colorimetry, and that most colours are reproducible with the proposed colorimetry. Viewers perceived conspicuous visual differences between the wide-gamut images and ones the gamut of which is limited to that of HDTV and felt that wide-gamut images appeared closer to the real objects.

3.2.7   Colour encoding

3.2.7.1 General

The use of a colour coding in transmission that is different from the one in acquisition and display, e.g. Y′Pb′Pr′, can reduce the transmission bandwidth while maintaining the luminance resolution that the human visual system (HVS) is sensitive to. The suitability of this method for UHDTV should be examined as follows.

−   What kind of primaries are suitable?

−   What are the ratios for spatial sampling between those primaries, e.g. the ratio between luminance and colour-difference?

−   What is the order of the transform matrix and non-linear conversion (Gamma correction)?

Conversion from RGB to luminance and colour difference signals was originally introduced when the colour television was developed. This was done while maintaining compatibility with black and white television. Later, it was used in the analogue component interface in a studio environment since it is less affected by error and noise. Then, when the interface was digitalized, the digital component of the signal format (Recommendation ITU-R BT.601) was standardized. The chroma sub-sample was introduced to reduce the total bit rate of the interface while taking advantage of human visual system characteristics. The major issue in the study was maintaining picture quality, when applied to chroma key processing in particular [13]. Those studies are described in detail in Reports ITU-R BT.629 Digital coding colour television signals and BT.962 The filtering, sampling and multiplexing for digital encoding of colour television signals. Based on broad studies including those ones, "4:2:2" was standardized as the chroma sub-sample ratio for the digital component signal in Recommendation ITU-R BT.601. It was also used for HDTV and described in Recommendation ITU-R BT.709.

The chroma sub-sample ratio affects the required bit rate of digital interface the most. The serial digital interface (SDI) for HDTV started at 1.5 gigabits per second that was available at that time to transmit 4:2:2 signals [14][15]. Later, in response to the requirement for higher picture quality,

1.5 Gigabits-per-second dual-link system was standardized to transmit 4:4:4 signals [14][16]. Further, advances in technology have enabled 3 Gigabits-per-second SDI and a single link system to be standardized to transmit 4:4:4 signals [14][17][18].

Picture quality is a trade-off for the "cost" of the interface when the chroma sub-sample ratio varies. It is again required from the "definition" of UHDTV described in [Input, 20] that UHDTV should have picture quality no worse than that of HDTV. Higher picture quality is desirable even in a simple post-production[2].

The interface seems slightly different from that used in HDTV where 1.5 Gigabits per second was the limit although it is true that lower bit rate is always preferable. This is because optical transmission technologies are going forward in the field of network technology, and those technologies are supposed to be applicable to the UHDTV interface. An optical transmission system was recently developed for the interface of UHDTV studio equipment [19]. It carries 72 Gigabits-per-second data that correspond to a 7680 x 4320 pixel-, 60Hz-, 12-bit-, 4:4:4- signal.

The discussions related to this section during study period 2008-2011 are as follows.

It was agreed that the answer to the second question (chroma sub-sample ratio) is that 2:1 on one-dimension is acceptable as the past experimental results show. The results are reflected to the parameter values of 4:2:2 and 4:2:0.

The third question means solving the so-called "constant luminance issue". It was agreed that the gamma correction is done after the transform matrix as the Y' equation in the PDNR shows. However, it is recognized that this is not independent from the first question.

Colour difference equations R'-Y' and B'-Y' are mentioned in the PDNR as the answer to the first question. These equations are slightly different from those of conventional colour difference signals as the component Y' changed from the conventional one. More specifically, the Y' in PDNR is not a linear combination of gamma-corrected R, G, B signals while the one for conventional systems is. The Y in PDNR is produced by combining linear R, G, B signals to maintain the constant luminance information. This fact brings about some differences in the output of some signal processing from that of R'G'B' case when the signals in PDNR are directly fed to it, i.e. without decoding to R'G'B' signals.

Note: Technical information relating ongoing discussion (as of September 2011) on colour encoding is tentatively attached as Attachment 1 through 6.

3.2.7.2 Constant luminance approach

The colour encoding process defined in the Recommendation ITU-R BT.709 was actually proposed to comprise the least computational steps, i.e., two encoding steps ($RGB \rightarrow R'G'B' \rightarrow Y'C'_BC'_R$) and one decoding step ($Y'C'_BC'_R \rightarrow R'G'B'$). The nonlinear $R'G'B'$ signals in the decoding step could be used without further transformation as an input to compensate the intrinsic nonlinear property of a CRT receiver. The latter in the encoding steps was then formulated to have the symmetric inverse structure against the decoding step. As a result, the luma Y' should have been formed by summing nonlinear $R'G'B'$ signals although these signals are not linearly proportional to the luminance (cd/m^2) information of a captured original scene. This fact induced crosstalk between Y' and $C'_BC'_R$ signals: the luma Y' signal includes chrominance information and vice versa. When colour-difference $C'_BC'_R$ signals are manipulated by sub-sampling them, the luminance information can also be varied because of the crosstalk property.

---

[2] The concept of simple/complex post-production is described in Recommendation ITU-R BT.1662 for LSDI application.

What can then be a criterion to choose an appropriate colour encoding method (video signal format) in the viewpoint of colour changes occurring after the sub-sampling process? If any colour stimuli are sub-sampled, the colour appearance on the resultant stimulus is perceived differently from that on each of the original stimuli to be sub-sampled. There is not a video signal set that can preclude this colour appearance change. Therefore, it will be the most appropriate to choose the video signal set providing fewer occasions where significant colour changes from the original due to the sub-sampling process are perceived. The image quality degradation can then be minimized although the original image data size is decreased by down-sampling chrominance signals. In conclusion, it will be reasonable to develop a new video signal format for UHDTV system offering fewer significant colour-variation cases after the sub-sampling process compared with that for HDTV system (Rec. ITU-R BT.709). This can be achieved by devising a new colour encoding method to be able to generate more orthogonal luma and colour-difference components than the conventional method of Rec. ITU-R BT.709.

UHDTV systems have been developing to offer viewers enhanced visual experiences compared to HDTV systems. One of key development factors for UHDTV is the sense of presence that viewers are aware of being there. To realize this requirement, large flat panel displays, currently using LCD technology, with higher spatial resolution (e.g., 3840×2160 and 7680×4320) than 1920×1080 for HDTV are under development mainly for broadcasting applications. The transition from HDTV to 3840 × 2160 UHDTV would imply that four times more pixels need to be transmitted to users, resulting in a four-fold increase in uncompressed bit rate. Today, each HDTV broadcast service uses all or a major part of a broadcast emission channel. A four-fold increase in uncompressed bit rate would thus be well beyond the capabilities of the source coding and channel coding methods currently used for HDTV broadcasting. In other words, the emission of a 3840 × 2160 UHDTV broadcasting service would require bit rates that exceed the capacity of a current terrestrial television channel." This fact suggests that improvement in the current coding technology and new video signals contributing to reducing bit rates are required for UHDTV systems to be used in broadcasting programme production/transmission pipe lines.

3.2.8 Bit depth

Bit depths of 8 and 10 have been empirically used for digital video signals of television applications. There is a related description in an early version of Rec. ITU-R BT.709 as follows.

Studies are continuing in this area. There is agreement that at least 8 bits are required for R, G, B, Y, C1, C2 and that 10 bits will be required for some applications. Therefore, both 8-bit and 10-bit representations are required.

The later version of Rec. ITU-R BT.709 recommends 8 and 10 as the bit depths, as is well known.

Rec. ITU-R BT.1769, Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange, recommends bit depths of 10 and 12, in consideration of emerging requirements for higher picture quality.

Let us briefly consider what the bit depth should be for UHDTV. The quality requirement is that the discontinuity in tone reproduction should be below a certain criterion. The most demanding criterion would be the perceptible limit, and it depends on the image content and viewing conditions.

The contrast sensitivity function (CSF) of the human visual system (HVS) relates to the ability to detect discontinuities in tone reproduction. The Weber-Fechner law (Eq. 2) is a classical representation of contrast sensitivity.

$$\text{Minimum detectable contrast: } \Delta L / L = const. \tag{2}$$

where,

L: luminance

ΔL : minimum detectable difference in luminance

According to this law, the minimum detectable contrast, i.e., the reciprocal of contrast sensitivity, is constant regardless luminance. It is believed that the ratio is between 1/50 and 1/100. However, it increases below and above certain luminances. Figure 23 shows measured data taken from Schreiber's book [20].

The actual CSF varies with the screen luminance, field of view, spatial frequency of the image, etc. Barten's model [21] takes such conditions into account. Figure 23 also shows the minimum detectable contrast calculated according to Barten's model with the assumed conditions[3] for UHDTV applications. Note that the minimum detectable contrast is calculated by using the following equation.

$$\text{Minimum detectable contrast} = 1/\text{CSF}_{\text{Barten}} \times 2 \times 1/1.27 \tag{3}$$

Here, 2 is used for the conversion from modulation to contrast and 1/1.27 is used for the conversion from sinusoidal to rectangular waves [22].

FIGURE 23

Examples of HVS minimum detectable contrast characteristics

---
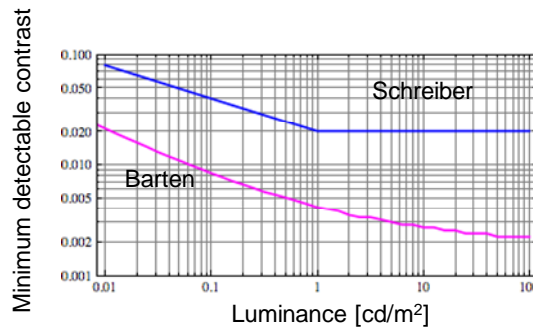
[3]   The following constants were used:

$k = 3.0$, $T = 0.1 \sec$, $\ddot{} = 0.03$, $\sigma_0 = 0.5 \arcmin$, $X_{\max} = 12°$, $\Phi_0 = 3 \times 10^{-8} \sec \deg^2$, $C_{ab} = 0.08 \arcmin/\text{mm}$, $N_{\max} = 15$ cycles , $u_0 = 7$ cycles/deg

where:

k : signal-to-noise ratio at 50% detection probability

T : integration time of the eye

$\ddot{}$: quantum efficiency of the eye

σ_0 : standard deviation of the optical line-spread function

$X_{\max}$: maximum integration area in the x direction

Φ_0 : spectral density of the neural noise

C_{ab} : aberration constant of the eye lens

$N_{\max}$: maximum number of integration cycles, and u_0 is the spatial frequency limit of the lateral inhibition process.

The angular size of the object X_0 is $60 \deg$, and the photon conversion factor p is $1.2274 \cdot 10^6 \text{photons/sec/deg}^2/\text{Td}$.

The bit depth and coding function should be determined by taking HVS characteristics into account so that the discontinuity in tone reproduction will be below a certain criterion. More specifically, the step (reproducible contrast) in the displayed luminance between consecutive codes of the coding function should be lower than the criterion, e.g. the minimum detectable contrast. Figure 24 shows the reproducible contrast of log and power coding functions. These functions are in fact the display EOTFs, i.e., the inverses of the coding functions, rather the coding functions themselves. The reproducible contrast is calculated with Eq. 2. The black level for the log function is set to 10E-6 of the white level.

FIGURE 24

Reproducible contrast comparison of log and power functions

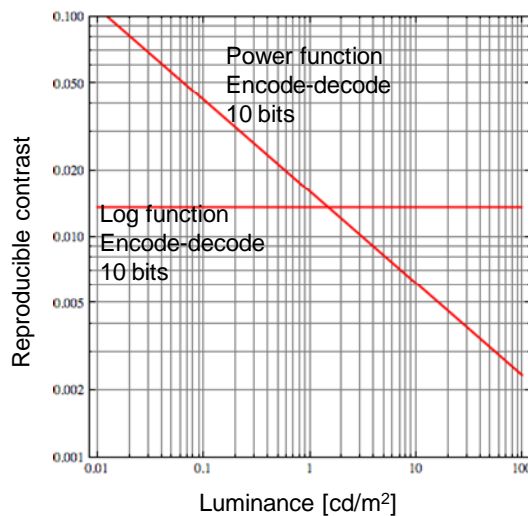The figure shows that the log function allots relatively more codes to the low luminance region, whereas the power function allots more to the high luminance region. The minimum detectable contrast increases in the low luminance region, and the luminance of reproduced video signals generally distributes more in the high luminance region. Hence, the power function is preferable to a log function.

Reference [22] has derived the optimal coefficient and bit depth for digital cinema applications. The criterion is the CSF calculated with Barten's model. A coding function with a derived coefficient of 2.6 and a bit depth of 12 ensures that the reproducible contrast of the function is less than the minimum detectable contrast calculated from the CSF of the HVS. The calculation was done for the luminance range of cinema applications, i.e., approximately 50 cd/m^2.

A coefficient of 2.4 may be preferable for UHDTV because the current CRT reference monitor has characteristics that would be close to those with a power function with a coefficient of 2.4, and consequently the current operational practice would be preserved. It should be noted that the coding function for the signal format is determined from the assumed EOTF of reference display.

Now let us turn to the bit depth. Figure 25 shows the reproducible contrast characteristics of power functions with a coefficient of 2.4 and 10- and 12-bit precisions. Three minimum detectable contrast characteristics mentioned above are also plotted in the figure for the sake of comparison. They are Schreiber's data, Barten's model for a flat signal, and the one for a ramp signal [22]. If we take Schreiber's data as a criterion, 10 bits is sufficient. If we take the Barten's model for rectangular signal even 12 bits is not sufficient for some ranges of luminance. However, when a natural scene is reproduced, a typical tone jump typically happens in parts where the luminance gradually changes.

Therefore, a ramp signal is more appropriate than a flat signal. Taking those factors into account, Fig. 25 shows that 12 bits is sufficient to cause no tone jump artefact.

FIGURE 25

**Comparison of reproducible contrasts of non-linear coding functions and HVS
minimum detectable contrast characteristics**
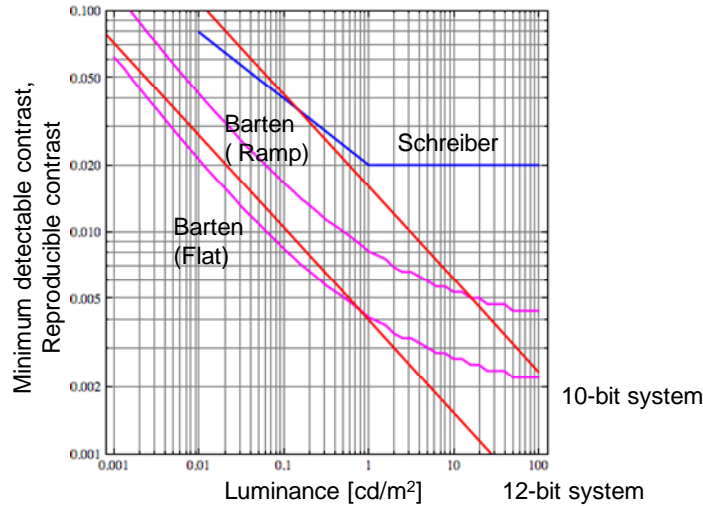
A wider colour gamut also expands the distance between consecutive codes given the same bit depth. Figure 26 shows the histogram of colour differences ΔE_{ab*} between all consecutive 8-bit codes for Rec. ITU-R BT.709 primaries and the new primaries proposed for UHDTV (Set b in Table 4 of [Input, 19]). The mean and maximum values are listed in Table 6. The mean of ΔE_{ab*} for UHDTV is 35% larger than that of HDTV. This means that 1bit more of precision is sufficient to make the ΔE_{ab*} for UHDTV equal to or less than that for HDTV. The maximum value for UHDTV will be below 1 if 10-bit coding is employed.

FIGURE 26

Histogram of colour errors in 8-bit coding
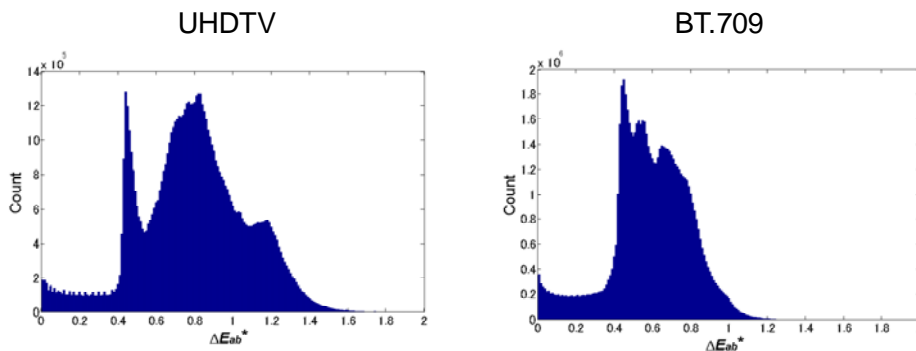
TABLE 6

Mean and Max of colour errors in 8-bit coding

Colorimetry	Mean	Max
Rec. ITU-R BT.709	0.58	1.45
Proposed for UHDTV	0.78	2.05

References

[1]   Mark Schubin, "Searching for the Perfect Aspect Ratio", SMPTE J., vol. 105, no. 8, pp. 460-479, 1996.

[2]   M. Emoto and M. Sugawara, "Flicker perceptions for wide-field-of-view and hold-type image presentations," IDW '09, pp. 1233-1234, 2009.

[3]   M. Sugawara, K. Omura, M. Emoto, and Y. Nojiri, "Temporal Sampling Parameters and Motion Portrayal of Television," SID 09 DIGEST, pp. 1200-1203, 2009.

[4]   M. Emoto, and Y. Matsuo, "Frame Frequency for Wide Field of View Video System - Dependency of Dynamic Visual Acuity on Field of View -," Annual Conv. of ITE, 2008.

[5]   M. Emoto *et al.*, "Improvement in Moving Picture Quality by High Frame Rate," The Journal of the ITE, Vol. 65, No. 8, pp. 1208-1214, 2011.

[6]   R. W. G. Hunt, The Reproduction of Colour, Sixth Edition: John Wiley & Sons, 2004.

[7]   Samsung Electronics Products produced in 2007 – 2009.

[8]   Digital Cinema System Specification, ver.1.0, Digital Cinema Initiatives, LLC Member Representatives Committee, 2005.

[9]   Recommendation ITU-R BT. 709-5, "Parameter values for the HDTV standards for production and international programme exchange", ITU, Geneva, Switzerland, 2002.

[10]   Adobe RGB® Color Image Encoding, Adobe Systems Incorporated. 2005.

[11]   SMPTE 170M-2004, "Composite analog video signal – NTSC for studio applications", 2004.

[12]   M. Kuwata, *et al.*, "A 65-in. slim (255-mm depth) laser TV with wide-angle projection optical system," JSID 17(11), pp. 875-882, 2009.

[13]   S. Baron and D. Wood, "The Foundation of Digital Television: The Origins of the 4:2:2 Component Digital Television Standard," SMPTE Motion Imaging Journal, pp. 327-334, September 2005.

[14]   Recommendation ITU-R BT.1120-7 Digital interfaces for HDTV studio signals, December 2007.

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

[16]   SMPTE 372M-2009 Dual Link 1.5 Gb/s Digital Interface for 1920 x 1080 and 2048 x 1080 Picture Formats.

[17]   SMPTE 424M-2006 Television - 3 Gb/s Signal/Data Serial Interface.

[18]   SMPTE 425-2008 3 Gb/s Signal/Data Serial Interface - Source Image Format Mapping.

[19]   M. Nakamura, T. Nakatogawa, and K. Oyamada, "Development of Optical Interface for Full-Resolution Super Hi-Vision Production Systems", ITE Tech, Rep., Vol. 33, No. 32, pp.5-8, July 2009.

[20]   W. F. Schreiber, Fundamentals of Electronic Imaging Systems, Third Edition: Springer-Verlag, 1992.

[21]   P. G. J. Barten, Contrast Sensitivity of the Human Eye and its Effects on Image Quality: SPIE PRESS, 1999.

[22]   M. Cowan, G. Kennel, T. Maier, and B. Walker, "Constant Sensitivity Experiment to Determine the Bit Depth for Digital Cinema," SMPTE Motion Imaging Journal, vol. 2004, pp. 281-292, 2004.

APPENDIX

Input documents to WP 6C or SG 6 related to UHDTV[4]

Input Ref.	Year	Document No.	Title	Submitter
1	2008	6C/39	Proposed draft new Questions ultra high definition television (UHDTV)	Japan
2		6C/40	Report on recent progress of technologies that are applicable to expanded hierarchy of LSDI, EHRI and UHDTV	Japan
3		6/62	Draft revision of Question ITU-R 40/6 - Extremely high-resolution imagery	WP 6C
4		6C/115	Proposed framework of work plan for the study on ultra high definition television	Japan
5		6/92	Research and development of Ultra High Definition Television (UHDTV)	EBU, BBC, RAI, NHK
6		6C/133, Annex 16	Framework of work plan for the study on ultra high definition television	Chairman, WP 6C
7		6C/133, Annex 18	Decision - Appointment of a Rapporteur Group - Commencement of study on UHDTV	Chairman, WP 6C
8	2009	6C/174	Report on the work plan and study of UHDTV	Rapporteur Group/UHDTV
9		6C/176	Tentative forecast for the applications and the time frame of UHDTV broadcasting services	CBS
10		6C/195	Proposed preliminary draft new Recommendation - Parameter values for extremely high-resolution imagery for production and international exchange of programmes	Korea
11		6C/TEMP/130	Proposed definition of UHDTV	WP 6C
12		6C/210, Annex 9	Working document towards a preliminary draft new Recommendation – Parameter values for extremely high-resolution imagery for production and international programme exchange	Chairman, WP 6C
13		6C/210, Annex 14	Working document for the study on UHDTV	Chairman, WP 6C
14		6C/210, Annex 20	Continuation of the Rapporteur Group - Study on UHDTV	Chairman, WP 6C
15		6C/246	Report on the study of UHDTV	Rapporteur Group/UHDTV

---

4   Documents related to audio are not included.

Input Ref.	Year	Document No.	Title	Submitter
16		6C/255	Proposed colorimetry for UHDTV	Japan
17		6C/257	Performance evaluation of compression coding for expanded hierarchy of LSDI and EHRI signals	Japan
18		6C/277	Proposed preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - Parameter values for UHDTV systems for production and international programme exchange	Korea
19		6C/287, Annex 9	Working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV]	Chairman WP 6C
20		6C/287, Annex 13	Studies relating to UHDTV	Chairman WP6C
21		6C/287, Annex 25	list of rapporteurs and rapporteur groups	Chairman WP6C
22	2010	6C/319	Report on the study of UHDTV	Rapporteur Group/UHDTV
23		6C/329Rev.1	Technical modification to WD-PDNR ITU-R BT.[IMAGE-UHDTV]	Korea
24		6C/330	Examination of two sets of new RGB primaries proposed in WD-PDNR ITU-R BT.[IMAGE-UHDTV]	Korea
25		6C/333	Proposed modification to working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV]	Japan
26		6C/353	Liaison statement on draft Recommendation J.LSSYS: Real-time transmission system of exlsdi signals under spatial image segmentation for parallel processing	ITU-T Study Group 9
27		6C/388	Work plan towards a new Recommendation ITU-R BT.[IMAGE-UHDTV]	Korea
28		6C/401	Capture and display experiments of wide colour gamut images	Japan
29		6C/404	Update of compression coding performance report for expanded hierarchy of LSDI and EHRI signals	Japan
30		6C/405	Proposed preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - UHDTV system parameters for production and international programme exchange	Japan
31		6C/407	PROPOSED DRAFT NEW RECOMMENDATION ITU-R BT.[COLOUR] Approaches of extended colour representation for television image systems	Japan
32		6C/415, Annex 7	Working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - Parameter values for UHDTV systems for production and international programme exchange	Chairman WP6C
33		6C415, Annex 18	Working document toward draft new Report ITU-R BT.[UHDTV] - The present state of ultra high definition television	Chairman WP6C

Input Ref.	Year	Document No.	Title	Submitter
34		6C415, Annex 19	Work plan for the study of the baseband image format for UHDTV systems	Chairman WP6C
35	2011	6C/425	Liaison statement - Real-time transmission system of exLSDI signals under spatial image segmentation for parallel processing	ITU-T SG 9
36		6C/437	Proposed preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - Parameter values for UHDTV systems for production and international programme exchange	Korea
37		6C/455	Consideration of luminance and colour-difference signal formats for UDTV image parameters	Japan
38		6C/456	Proposed modification to working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - Inclusion of a higher frame frequency - UDTV system parameters for production and international programme exchange	Japan
39		4C/461	Some considerations on broadcasting applications for UHDTV	CBS, Inc.
40		6C/466	Current status of discussions concerning colour equations for UHDTV systems	Chairman WP6C
41		6C/490, Annex 5	Preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] –Parameter values for UHDTV systems for production and international programme exchange	Chairman WP6C
42		6C/490, Annex 14	Preliminary draft new Report ITU-R BT.[UHDTV] – The present state of ultra-high definition television	Chairman WP6C
43		6C/490, Annex 17	Work plan for the study of the baseband image format for UHDTV systems	Chairman WP6C
44		6C/496	International Organization for Standardization - ISO/IEC JTC1/SC29/WG11 - Coding of moving pictures and audio – Liaison Statement on High Efficiency Video Coding (HEVC)	ISO
45		6C/501	Comment on PDN Recommendation ITU-R BT.[IMAGE-UHDTV] - Parameter values for UHDTV systems for production and international programme exchange	Italy
46		6C/504	Draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - Parameter values for UHDTV systems for production and international programme exchange	United States of America
47		6C/519	Recent study on high frame frequency television to support parameter values described in PDNR ITU-R BT.[IMAGE-UHDTV] – UHDTV system parameters for production and international programme exchange	Japan
48		6C/520	Proposed modifications to preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] for a draft new Recommendation - Parameter values for UHDTV systems for production and international programme exchange	Japan

Input Ref.	Year	Document No.	Title	Submitter
49		6C/528	Performance evaluation of compression coding for expanded hierarchy of LSDI and EHRI signals	Japan
50		6C/535	Report on the study of UHDTV	Rapporteur Group/UHDTV
51		6C/553	Proposed draft new Recommendation ITU-R BT.[IMAGE-UHDTV] - Parameter values for UHDTV systems for production and international programme exchange	Korea
			Other input documents to WP 6C	
52	2008	6C/114	Proposal of required electro-optical transfer function for a FPD as a master monitor	Japan
53	2009	6C/170	A perception based image representation for inclusion into the draft modification of Recommendation ITU-R BT.1361	United States of America

ATTACHMENT 1

The forward and inverse computational processes for the two types of signal format proposed in [Input, 32]

At the previous meeting in October 2010, WD-PDNR ITU-R BT.[IMAGE-UHDTV] "Parameter values for UHDTV systems for production and international programme exchange" [Input, 32] was produced. Two types of signal format that defines a colour encoding method for producing video signals were introduced in that WD-PDNR document. These two signal formats were originally suggested by Korea and Japan. The forward and inverse computational processes for the two signal formats are described in Table 7.

Table 7

The forward and inverse computational processes for the two signal formats suggested by Korea and Japan

(a) Forward computational processes

forward	Korea (RGB to $AC_{YB}C_{RG}$)	Japan (RGB to $Y'C_B'C_R'$)
[Step 1]	$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} 0.66930 & 0.15257 & 0.17824 \\ 0.26240 & 0.67850 & 0.0592 \\ 0.00009 & 0.02608 & 0.97374 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix}$$ where the $X_rY_rZ_r$ are normalizedtristimulus components using the $X_nY_nZ_n$ values of D65 reference-white point (X_n=0.9504, Y_n=1, and Z_n=1.0889).	$$Y = \begin{bmatrix} 0.2627 & 0.6780 & 0.0593 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix}$$
[Step 2]	$$E' = E^{0.45}$$ where E is each of normalized $X_rY_rZ_r$ components, and E' is each of nonlinear normalized $X'_rY'_rZ'_r$ components.	$$E' = \begin{cases} 4.5E, & 0 \le E < \beta \\ \alpha E^{0.45} - (\alpha - 1), & \beta \le E \le 1 \end{cases}$$ where E is Y or voltage normalized by the reference white level and proportional to the implicit light intensity that would be detected with a reference camera colour channel R, G, B; E' is the resulting nonlinear signal. α = 1.099 and β = 0.018 for 10-bit system α = 1.0993 and β = 0.0181 for 12-bit system
[Step 3]	$$\begin{bmatrix} A \\ C_{YB} \\ C_{RG} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -0.22865 & -0.12936 & 0.35801 \\ 0.64759 & -0.64719 & -0.0004 \end{bmatrix}\begin{bmatrix} X'_r \\ Y'_r \\ Z'_r \end{bmatrix}$$	$$\begin{bmatrix} Y' \\ C_B' \\ C_R' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/1.9403 & 1/1.9403 & 0 \\ -1/1.7182 & 0 & 1/1.7182 \end{bmatrix}\begin{bmatrix} Y' \\ B' \\ R' \end{bmatrix}$$

(b) Inverse computational processes

inverse	Korea ($AC_{YB}C_{RG}$ to RGB)	Japan ($Y'C_B'C_R'$ to RGB)
[Step 1]	$$\begin{bmatrix} X'_r \\ Y'_r \\ Z'_r \end{bmatrix} = \begin{bmatrix} 1 & 0.00173 & 1.54480 \\ 1 & 0 & 0 \\ 1 & 2.79432 & 0.98661 \end{bmatrix}\begin{bmatrix} A \\ C_{YB} \\ C_{RG} \end{bmatrix}$$	$$\begin{bmatrix} Y' \\ B' \\ R' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1.9403 & 0 \\ 1 & 0 & 1.7182 \end{bmatrix}\begin{bmatrix} Y' \\ C_B' \\ C_R' \end{bmatrix}$$
[Step 2]	$$E = E'^{(1/0.45)}$$ where E' is each of nonlinear normalized $X'_r Y'_r Z'_r$ components and E is each of normalized $X_r Y_r Z_r$ components..	$$E = \begin{cases} E'/4.5, & 0 \le E \le \beta \\ [(E'+(\alpha-1))/\alpha]^{(1/0.45)}, & \beta \langle E \le 1 \end{cases}$$ where E' is Y' or G' or B' nonlinear signal. α = 1.099 and β = 0.081 for 10-bit system α = 1.0993 and β = 0.0814 for 12-bit system
[Step 3]	$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.63401 & -0.35677 & -0.27741 \\ -0.63340 & 1.61559 & 0.01772 \\ 0.01681 & -0.04324 & 1.02652 \end{bmatrix}\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}$$	$$G = \begin{bmatrix} 1/0.6780 & -0.3875 & -0.0875 \end{bmatrix}\begin{bmatrix} Y \\ R \\ B \end{bmatrix}$$

ATTACHMENT 2

The evaluation results that were considered to devise a compromise between two previously suggested formats in [Input, 32]

1   The influence of different colour-encoding methods on the objective and subjective properties of the reconstructed images after subsampling colour difference components

It had already mentioned and demonstrated in [Input, 18] that the colour encoding method defined in Recommendation ITU-R BT.709 cannot preserve the luminance information of an original scene and so generate video signals having crosstalk problems. In order to solve these problems of the conventional colour encoding method, Korea firstly proposed a new colour encoding method in [Input, 18] and then Japan also suggested that in [Input, 30]. Both proposals adopted the same method in the production of a luma video signal, however, suggested different ways in the production of colour-difference video signals. These Korean and Japanese proposals summarised in [Input, 32] are re-introduced in Tables 8(a) and 8(b) respectively. The luminance video signals indicated by '$Y_r(A)$' in Table 8(a) and by '$Y(Y')$' in Table 8(b) are actually identical and are generated by adding linear *RGB* values multiplied by the coefficients necessary to form D65 white point. On the contrary, the colour-difference video signals represented by 'C_{YB} and C_{RG}' in Table 8(a) and by 'C_B' and C_R''' in Table 8(b) are produced differently.

TABLE 8

Colour encoding methods suggested by

(a) Korea

	RGB signals	Luma (*A*) and colour-difference (C_{YB} and C_{RG}) signals
[Step 1]	$T' = T^{0.45}$ where *T* is each of RGB tristimulus values in the range of $0-1$ and T' is each of nonlinear RGB tristimulus values.	$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} 1/0.9504 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/1.0889 \end{bmatrix}\begin{bmatrix} 0.63606 & 0.14498 & 0.16936 \\ 0.26237 & 0.67847 & 0.05916 \\ 0.00009 & 0.02837 & 1.06034 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix}$$ where the XYZ values are normalized using the $X_nY_nZ_n$ values of D65 reference-white point (X_n=0.9504, Y_n=1, and Z_n=1.0889).
[Step 2]		$$E' = E^{0.45}$$ where *E* is each of normalized $X_rY_rZ_r$ encoding components, and E' is each of nonlinear normalized $X'_rY'_rZ'_r$ encoding components.
[Step 3]		$$\begin{bmatrix} A \\ C_{YB} \\ C_{RG} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -0.22865 & -0.12936 & 0.35801 \\ 0.64759 & -0.64719 & -0.0004 \end{bmatrix}\begin{bmatrix} X'_r \\ Y'_r \\ Z'_r \end{bmatrix}$$

(b) Japan

Parameter	Values	
Signal format	R',G',B'	Y' (Luma), C_B' and C_R' (colour differences)
Nonlinear transfer characteristics	$$E' = \begin{cases} 4.5\,E, & 0 \le E < \beta \\ \alpha E^{0.45} - (\alpha - 1), & \beta \le E \le 1 \end{cases}$$ where E is voltage normalized by the reference white level and proportional to the implicit light intensity that would be detected with a reference camera colour channel R, G, B; E' is the resulting nonlinear signal. $\alpha = 1.099$ and $\beta = 0.018$ for 10-bit system $\alpha = 1.0993$ and $\beta = 0.0181$ for 12-bit system	
Derivation of luminance signal Y'	$Y' = (0.2627R + 0.6780G + 0.0593B)'$ The nonlinear transfer characteristics are specified in the above row.	
Derivation of colour difference signals C_B', C_R'	$C_B{}' = \dfrac{B' - Y'}{1.9403}$, $C_R{}' = \dfrac{R' - Y'}{1.7182}$	

The influence of the difference in the colour-difference signals for the two proposals expressed in Tables 8(a) and 8(b) were examined after subsampling colour-difference signals in the format of 4:2:0 and 4:1:0. Three still test images containing fine textures and sharp edges were chosen for examining the level of crosstalk occurring between the luma signal and each of two different colour-difference signal sets given in Tables 8(a) and 8(b). These test images are shown in Fig. 27. A PSNR (Peak Signal to Noise Ratio) lightness difference was chosen to assess the amount of crosstalk and CIELAB L^* was used to represent perceived lightness. Fig. 28 illustrates the computational procedure for the PSNR-lightness measure between an original image and its counterpart generated after 'only chroma subsampling' or 'subsampling and compression' process. To compare image quality between the original and its counterpart image on a monitor, sRGB transformation was used [23]. If the *RGB* data in an original image were encoded in the form of nonlinear Recommendation ITU-R BT.709 *R'G'B'*, the transformation given in Recommendation ITU-R BT.709 was used instead of the sRGB conversion. The CIELAB colour transformation was utilized in the calculation of lightness values [24].

FIGURE 27

Three still test images containing fine textures and sharp edges

Test Image 1	**Test Image 2**	**Test Image 3**

(a)	**(b)**	**(c)**
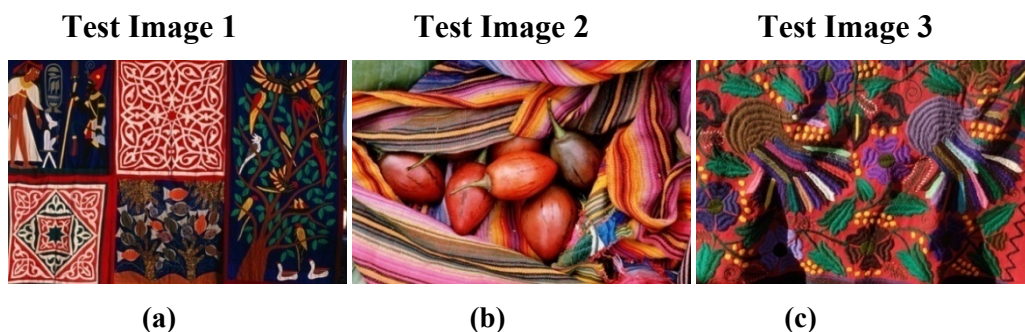
FIGURE 28

A computational procedure for PSNR-lightness between an original image and its chroma subsampled image (or its compressed image after subsampling chorma components)
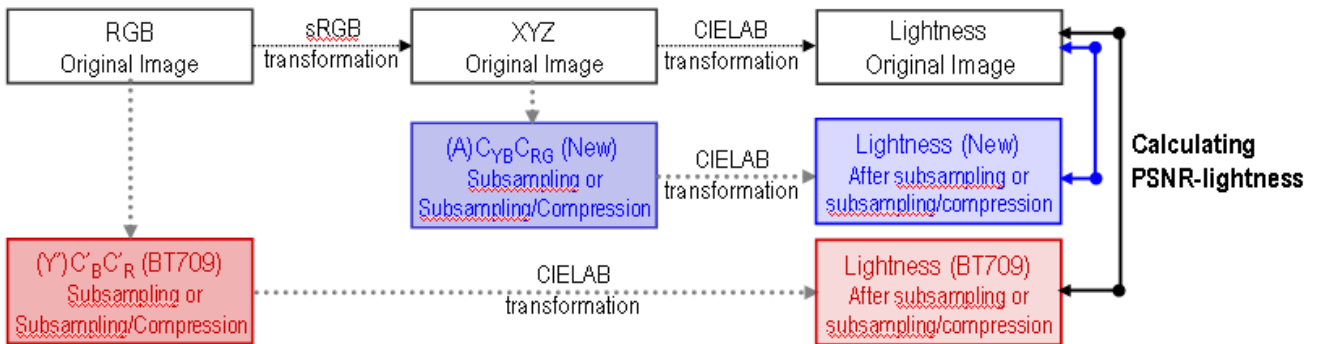
For the evaluation of the level of crosstalk, three colour-encoding methods were used to produce three sets of luma and colour-difference signals: Recommendation ITU-R BT.709, Korean proposal (see Table 8(a)) and Japanese proposal (see Table 8(b)). Figures 29(a) and 29(b) show the PSNR-lightness values calculated between the original images (see Fig. 27) and their upsampled (reconstructed) images after subsampling colour-difference components. The results obtained from the three encoding methods are represented differently: grey bars for Recommendation ITU-R BT.709, blue bars for the Korean proposal and red bars for the Japanese proposal.

FIGURE 29

The computed PSNR-lightness values between the original and its reconstructed image after subsampling colour-difference components using (a) 4:2:0 and (b) 4:1:0 formats

(a)   (b)
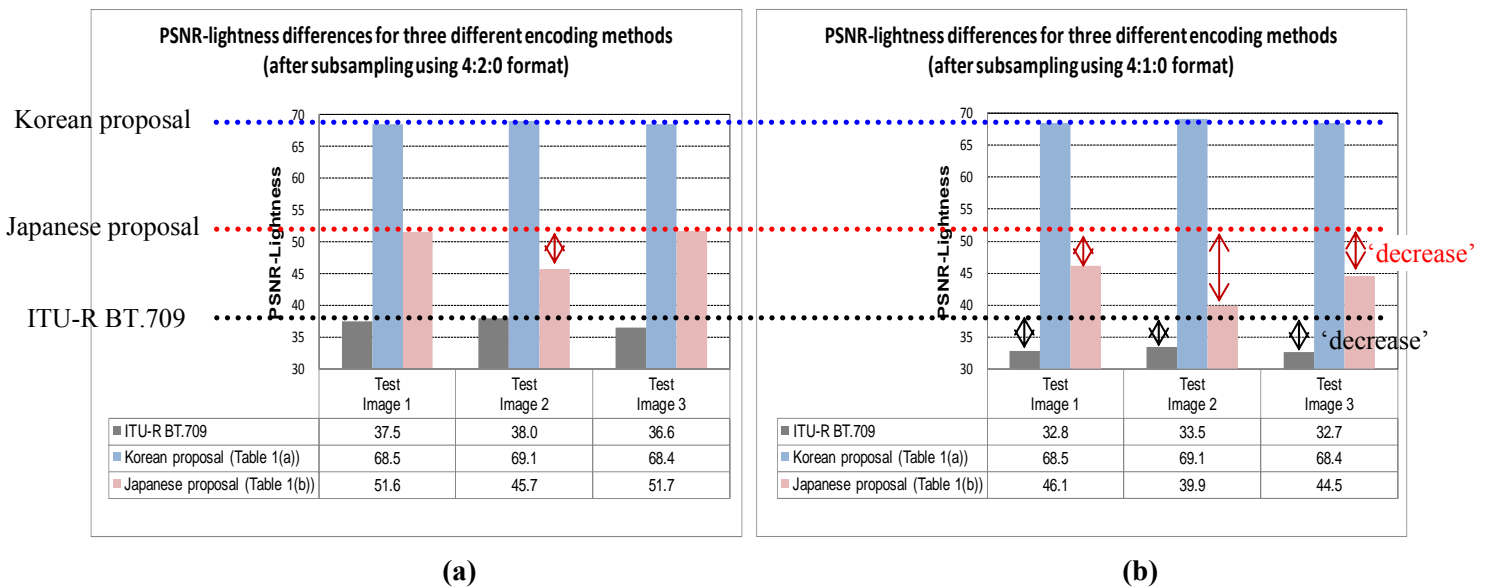
When the resulting PSNR-lightness values shown in Figs. 29(a) and 29(b) are compared for the three encoding methods, the following results are found.

1)   For both 4:2:0 (see Fig. 29(a)) and 4:1:0 (see Fig. 29(b)) formats, the largest PSNR-lightness values are obtained from the Korean proposal described in Table 8(a) and the

smallest values from Recommendation ITU-R BT.709. This result indicates that the luminance information of the original images shown in Fig. 27 is the most varied after subsampling colour-difference signals for the Recommendation ITU-R BT.709 whereas that is the least changed for the Korean proposal. The Japanese proposal explained in Table 8(b) produces middle PSNR-lightness values that are smaller than the Korean proposal but larger than Recommendation ITU-R BT.709.

2)   When a subsampling rate is increased from 4:2:0 (see Fig. 29(a)) to 4:1:0 (see Fig. 29(b)), a reduction in the PSNR-lightness values is observed for the Japanese proposal and the ITU-R BT.709. This tendency is indicated by red dotted line and arrows for the Japanese proposal and by black dotted line and arrows for Recommendation ITU-R BT.709. The PSNR-lightness values are however the same even in the case of raising subsampling rate for the Korean proposal that is shown using blue dotted line.

The Korean and Japanese proposals have the same luma signal, but different colour-difference signals. Both proposals show smaller lightness differences between the original images and their reconstructed images after subsampling colour-difference components than Recommendation ITU-R BT.709, suggesting both proposals are effective in solving the crosstalk problems that Recommendation ITU-R BT.709 has. However, the above two findings demonstrate that the Korean proposal can separate luminance and chrominance information more accurately than the Japanese proposal. In order to find out whether this deviation in terms of PSNR-lightness measure causes noticeable subjective quality difference, a couple of observers compared all the reconstructed images after subsampling colour-difference components for the original test images shown in Fig. 27. The images to be evaluated were divided into two groups according to the subsampling rate, 4:2:0 or 4:1:0, and presented on a monitor. There was not a significant quality difference between two sets of the reconstructed images that were produced using the Korean and Japanese proposals. Those formed using Recommendation ITU-R BT.709 were all judged to have more blurred appearance and less sharp edges than the Korean and Japanese proposals. Note that these subjective quality evaluations were conducted unofficially, therefore if it is necessary; an official subjective quality evaluation will have to be carried out by an independent body.

Amongst three colour-encoding methods to be examined, it can be said that the best one producing luminance and chrominance video signals containing the smallest crosstalk is Korean proposal when judging based on the comparison results (see Figs. 29(a) and 29(b)) using the objective measure of PSNR-lightness. The deviations in the PSNR-lightness measure between the Korean and Japanese proposals could not however be led to the perceived quality differences. This is probably due to the fact that two different images having PSNR-lightness greater than a certain value are very likely perceived similarly. Both Korean and Japanese proposals use the same method to create luma video signal, but not identical ways to produce colourdifference signals. Therefore, it can be thought that the luminance video signal plays more important role than the colourdifference signals in the separation of luminance and chrominance information. In conclusion, the encoding method to produce independent luma video signal should be adopted from the Korean and Japanese proposals for UHDTV baseband image format. This will result in an improved quality after subsampling colour-difference components compared with the conventional encoding method of Recommendation ITU-R BT.709.

2   The influence of different colour-encoding methods on an image compression in terms of objective measures

The influence of different encoding methods on the quality of compressed images and compression efficiency was investigated using two moving images (1920H ×1080V) shown in Figs 30(a) and 30(b) that were purchased from The Institute of Image Information and Television Engineers in Japan (http://www.ite.or.jp/shuppan/testchart_index.html#HI10). Actually, NHK Engineering Co.

produced these moving images. The test image seen in Fig. 30(a) includes many colourful objects while that seen in Fig. 30(b) appears to be achromatic, i.e., much less colourful scene. These two sets of test sequence were provided in TIFF format composed of nonlinear Recommendation ITU-R BT.709 *R′G′B′* for 59.94 Hz/interlace scanning. Only 61 frames in the original test sequences were used to analyse compression coding efficiency and the quality of the compressed images. The *R′G′B′* data in the original images were converted to the conventional video signals of Recommendation ITU-R BT.709 and the new video signals using the Korean and Japanese proposals that were explained in Tables 8(a) and 8(b). After subsampling and low-pass filtering colourdifference signals in the format of 4:2:2, the chroma-subsampled images were then compressed using the reference software JM12.0 [25-28]. The encoder configurations applied to compress the test sequences are as follows.

1)   High profile

2)   GOP structure: I-B-B-P-B-B-P

3)   Entropy coding mode: CABAC

4)   R-D optimization enabled

5)   Number of reference and search range: 1 and 64

6)   QP for I/P/B slice = 22/23/24, 27/28/29, 32/33/34, 37/38/39

FIGURE 30

Two moving test images

Test Sequence 1   **Test Sequence 2**

(a)   **(b)**
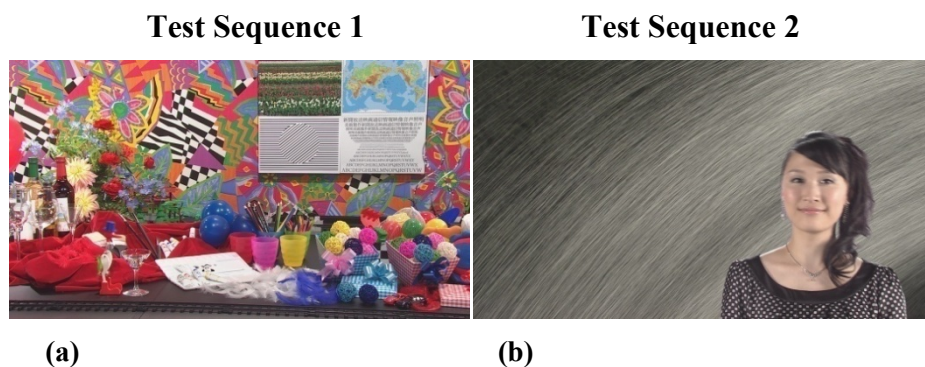
Different luma and colour-difference signals, which were produced from the conventional Recommendation ITU-R BT.709, Korean and Japanese encoding methods, were used on the compression process. Instead of PSNR-luma, the PSNR-lightness L^*, -chromaC^*, -hue *h* and -colour difference ΔE^*_{ab} were therefore used to measure objective quality differences arising between the original and its compressed images. The computational procedure for the PSNR-lightness was illustrated in Fig. 28 where the lightness L^* values were calculated from *XYZ* values by the CIELAB transformation. Not only L^* but also C^* and *h* can be obtained by the same conversion process. The colour difference can then be calculated using all of the L^*C^* *h* values. Both luma and colourdifference components are manipulated on the compression process whereas only colourdifference signals are altered on the colour-difference subsampling process. This fact indicates that both luminance and colour information in the original would be changed after the compression process. Thus, variations in the colour and hue domains as well as those in the lightness domain were calculated between the original and its compressed versions.

Figures 31(a) and 31(b) show the changes in the PSNR-lightness, -colour, -hue and -colour difference values against bit-rates (kbits/sec) at 59.94 Hz for the two test sequences seen in Figs 30(a) and 30(b). The BD-PSNR and BD-rate values were calculated between the data points of the Recommendation ITU-R BT.709 and those of each of other encoding methods plotted in Figs. 31(a)

and 31(b) and are introduced in Table 9. The mean of the BD-PSNR and BD-rate values were calculated against the very colourful test sequence 1 and the neutral test sequence 2 and are also described in Table 9. The effects in the change of video signals produced by different encoding methods on the quality of compressed images can be summarised in the following.

1)   The largest mean BD-rate (about 13 – 15 %) and the smallest mean BD-PSNR (about -0.3 – -0.4 dB) values in the chroma, hue and colour difference domains are seen for the Japanese proposal in Table 9, explaining that the Japanese proposal (see Table 8(b)) shows much worse coding performance than the Korean proposal. In Figs. 31(a) and 31(b), the pink lines indicating Japanese proposal in the PSNR-chroma, hue and colour difference domains are located at the lowest vertical places. These results suggest that approximately 13 – 15 % higher bit-rate is needed for the Japanese proposal than for the Recommendation ITU-R BT.709 supposing that the same PSNR values are achieved between the Japanese proposal and the Recommendation ITU-R BT.709's encoding method. These results can probably arise from the fact that the C_B' C_R' colour difference signals in the Japanese proposal are quantised only into part of the available range (16 – 240 using 8 bit-depth), i.e., 16 – 219 and 16 – 196 respectively. This type of quantisation can cause the chrominance information of an original image to be changed more than the quantisation using a whole available range. As a result, the chroma and hue characteristics in the compressed sequences can be much more deviated from those of the original sequences for the Japanese proposal than other colour-encoding methods using a whole quantised range.

2)   It was proved in the previous section that the Korean proposal (see Table 8(b)) could separate luminance and chrominance information most accurately amongst the colour-encoding methods to be evaluated, leading to luma and colour-difference video signals containing the least crosstalk. In Table 9, the Korean proposal having this property gave the smallest mean BD-rate (about -20 – -27 %) and the largest mean BD-PSNR (about 0.6 – 0.7 dB) values in the chroma, hue and colour difference domains that are contrary results to the Japanese proposal. In Figs 31(a) and 31(b), the blue lines indicating Korean proposal in the PSNR-chroma, hue and colour difference domains are located at the highest vertical places. These results suggest that approximately 20 – 27 % less bit-rate is required for the Korean proposal than for Recommendation ITU-R BT.709 supposing that the same PSNR values are achieved between the Korean proposal and Recommendation ITU-R BT.709's encoding method.

3)   In the lightness dimension, slightly better coding performance is obtained using the Japanese proposal than the Korean proposal. However, the differences in the BD-PSNR and BD-rate values between the two proposals in the lightness dimension are much smaller than those in the chroma, hue and colour difference dimensions: 5.6 ΔBD-rate and 0.3 ΔBD-PSNR in the lightness dimension, 39.8 ΔBD-rate and 1.1 ΔBD-PSNR in the chroma dimension, 41.7 ΔBD-rate and 0.9 ΔBD-PSNR in the hue dimension, and 33.9 ΔBD-rate and 0.9 ΔBD-PSNR in the colour difference dimension.

In conclusion, a compromise between the Korean and Japanese proposals in the aspect of BD-PSNR and BD-rate measures was devised. The same generation method as both Korean and Japanese proposals' was adopted for producing a luma video signal. The way to create the colourdifference signals was modified from the Japanese proposal (see Table 8(b)) so as to allow the colour-difference signals $C_B'C_R'$ to be quantised into a whole available range. The BD-PSNR and BD-rate values were calculated for this compromise and are shown in Table 9. The compromise method performed better than the Korean proposal in the lightness domain and also than the Japanese proposal in the chroma, hue and colour difference domains. Therefore, the compromise colour-encoding method is proposed for a new recommendation to be used for UHDTV program productions and international exchanges.

FIGURE 31

The PSNR-lightness, -chroma, -hue and -colour difference values against the bit-rates obtained after compressing 4:2:2 subsampled images for (a) Test Sequence 1 and (b) Test Sequence 2
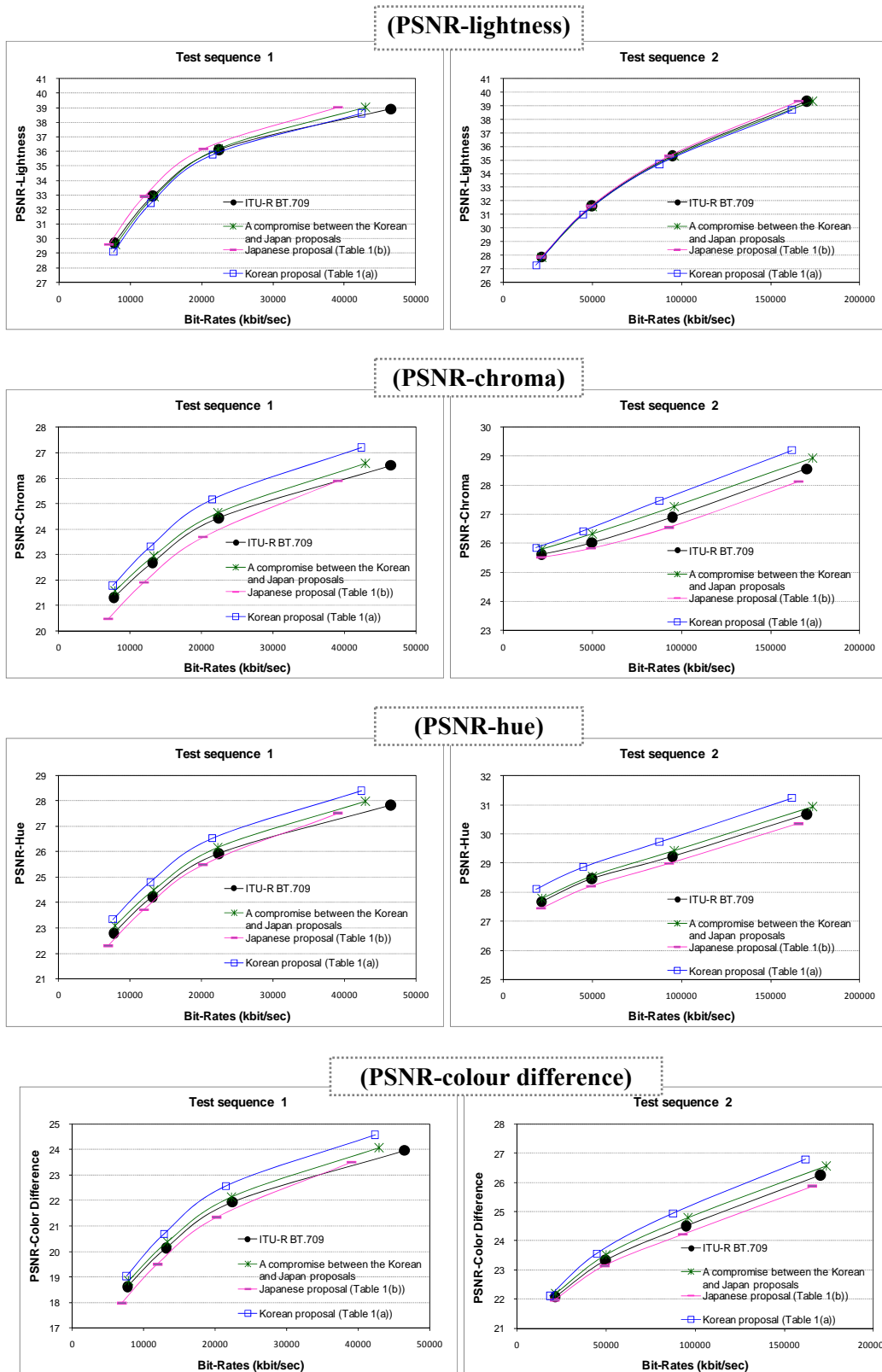
TABLE 9

The calculated results for the BD-PSNR and BD-rate measures that were obtained from each of three encoding methods (Korea, Japan and a compromise between them) in comparison with the Rec. ITU-R BT.709's encoding method

		Lightness		Chroma		Hue		Colour Difference	
		BD-rate	BD-PSNR	BD-rate	BD-PSNR	BD-rate	BD-PSNR	BD-rate	BD-PSNR
Test sequence 1	Korea	3.8	-0.2	-22.3	0.8	-21.0	0.7	-19.7	0.7
	Japan	-4.0	0.2	22.1	-0.6	11.9	-0.3	16.3	-0.5
	A compromise	-0.4	0.0	-7.2	0.2	-8.3	0.2	-7.0	0.2
Test sequence 2	Korea	1.9	-0.1	-26.5	0.6	-34.2	0.6	-21.1	0.5
	Japan	-1.5	0.1	8.7	-0.2	16.4	-0.2	10.8	-0.2
	A compromise	1.5	-0.1	-12.4	0.3	-9.6	0.1	-9.2	0.2
Mean of two test sequences	Korea[k]	2.9	-0.2	-24.4	0.7	-27.6	0.6	-20.4	0.6
	Japan[j]	-2.7	0.1	15.4	-0.4	14.1	-0.3	13.5	-0.3
	A compromise	0.6	0.0	-9.8	0.3	-9.0	0.2	-8.1	0.2
ΔBD-rate and ΔBD-PSNR values computed between the Korean[k] and Japanese[j] proposals using the mean results of two test sequences		5.6	0.3	39.8	1.1	41.7	0.9	33.9	0.9

NOTE – As the BD-rate value decreases, the target encoding method is more efficient than that of Recommendation ITU-R BT.709 in the aspect of a bit-rate reduction assuming that the same PSNR value between the target and Recommendation ITU-R BT.709's encoding methods is achieved. As the BD-PSNR value increases, better subjective quality is expected for the target encoding method than for that of Recommendation ITU-R BT.709 supposing the same bit-rate between the target and Recommendation ITU-R BT.709's encoding method.

3 Conclusions

[Input, 18] demonstrated that the luma and colour-difference video signals defined in the Recommendation ITU-R BT.709 were not independent signals, that is, the luma signal had chrominance information and the colour difference signals had luminance information. In addition, it was also shown that this crosstalk problem could cause the reconstructed images after subsampling the colour difference components to be appeared to have worse quality than the case using independent video signals.

Korea firstly proposed a new colour encoding method in [Input, 18] and then Japan also suggested that in [Input, 30] aiming at producing new video signals having insignificant crosstalk. In both proposals, the same method for producing a newlumavideo signal was suggested; however different ways for new colour-difference video signals were given.

In the present document, the evaluation results for Recommendation ITU-R BT.709, Korean and Japanese proposals were introduced. Image quality changes in the reconstructed images after subsampling colour difference components in the format of 4:2:0 and 4:1:0 were estimated using

objective and subjective measures. Those in the compressed images after subsampling and compression processes were also evaluated only using an objective measure. Key evaluation findings are summarised as follows.

− In the evaluation results only after the subsampling process:

  − the Korean proposal showed the best performance in terms of PSNR-lightness (CIELAB L*) representing the objective measure; however

  − both Korean and Japanese proposals provided quite similar subjective quality one another (but both looked better than the Rec. ITU-R BT.709).

− From the above results, it can be concluded that the luma video signal that is the same in both Korean and Japanese proposals plays more significant role than the colour difference signals that are not identical in both proposals. Therefore, the luma-signal producing method suggested by both proposals should be adopted for a new signal format for UHDTV systems.

− In the evaluation results after the subsampling and compression processes:

  − the Japanese proposal performed much worse than the Korean proposal by changing the chrominance information of original sequences in a very large amount (this is probably because of using only part of a whole available range in the quantisation of the colour difference signals for the Japanese proposal); whereas;

  − the Korean proposal performed slightly worse than the Japanese proposal in the aspect of variations of the luminance information of the original sequences.

− In conclusion, a compromise between the Korean and Japanese proposals was devised showing improved performance than the Korean proposal in the lightness domain than the Japanese proposal in the chroma, hue and colour difference domains after the compression process.

− The compromised colour-encoding method could produce the same luma signal as the Korean and Japanese proposals' and the modified colour-difference signals from the Japanese proposal. The modified colour-difference signals were derived in order to be quantised into a whole available range and have a centre quantised value for neutral colours.

− The compromised colour-encoding method was therefore suggested to be applied for the signal format of UHDTV systems.

References

[23]   IEC 61966-2-1, "Multimedia Systems and Equipment – Color Measurement and Management, Part 2-1: Color Management – Default RGB Color Space" (1999).

[24]   Supplement No. 2 of CIE Publication 15 (E-1.3.1) 1971,"Recommendation on Uniform Color Spaces, Color-Difference Equations, Psychometric Color Terms" (1978).

[25]   Recommendation ITU-T H.264 and ISO/IEC 14496-10, Advanced Video Coding for Generic Audiovisual Services, May 2003.

[26]   Recommended simulation common conditions for coding efficiency experiments, Document VCEG-AA10, ITU-T Study Group 16 Question 6, VCEG, ITU-T, October 2005.

[27]   Color format down-conversion for test sequence generation, MPEG2003/N6295, ISO/IEC JTC1/SC29/WG11, Waikoloa, December 2003.

[28]   Color format up-conversion for video display, MPEG2003/N6296, ISO/IEC JTC1/SC29/WG11, Waikoloa, December 2003.

ATTACHMENT 3

The scientific foundations applied to derive a new colour-encoding scheme suitable for UHDTV systems

1 A new colour-encoding scheme

Video signals are generally composed of two elements having luminance and chrominance information since Georges Valensi's suggestion in 1938. Recommendation ITU-R BT.709 defines an encoding scheme producing video signals corresponding to these two elements; luminance (Y') and colour-difference (C'_B and C'_R) signals. The spatial detail information in an image is mainly reproduced using a luminance channel, not a chrominance channel. Human eyes perceive more sensitively spatial changes in the luminance dimension than those in the chrominance dimension. Due to these two facts, only colour-difference signals are subsampled to reduce the image data to be transmitted whereas luma signals are maintained. However, if there is some luminance information in colour-difference signals, the luminance information can also be manipulated by subsampling the colour-difference values. As a result, the quality of a decoded image after the subsampling process is perceived to be lower than that of an original image, especially in the aspect of an accurate reproduction of edge and detail areas. This indicates that an efficient encoding scheme should enable luminance and chrominance information to be completely separated. In other words, there should be as small as possible crosstalk between luma and colour-difference channels. The level of crosstalk was firstly evaluated for Recommendation ITU-R BT.709 that is currently used to produce HDTV programmes. A new encoding scheme suitable for UHDTV programme production was then developed and aimed at creating video signals having insignificant crosstalk.

1.1 An investigation of crosstalk occurring between luma and colour-difference channels that are generated by the encoding scheme defined in Recommendation ITU-R BT.709

Recommendation ITU-R BT.709 produces luma and colour-difference signals according to the following workflow that is illustrated in Fig. 32. Linear RGB signals are converted onto nonlinear RGB signals using the opto-electronic transfer function having an exponent of 0.45 (i.e. gamma).

FIGURE 32

A block diagram illustrating the colour encoding scheme defined in Recommendation ITU-R BT.709

$$Y' = 0.2126R' + 0.7152G' + 0.0722B'$$
$$C'_B = (B' - Y')/1.8556$$
$$C'_R = (R' - Y')/1.5748$$
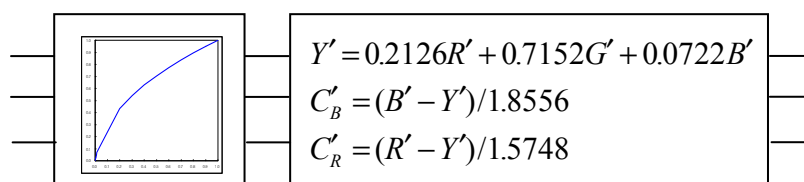
The luma component (Y') containing luminance information is computed from the weighted sum of nonlinear $R'G'B'$ signals. The different coefficient applied to each of $R'G'B'$ signals represents the relative luminance of the RGB primaries, which are defined in Recommendation ITU-R BT.709, in order to generate a reference white point of D65. In principle, these coefficients must be applied to each of linear RGB signals, not nonlinear signals, because linear RGB signals are proportional to
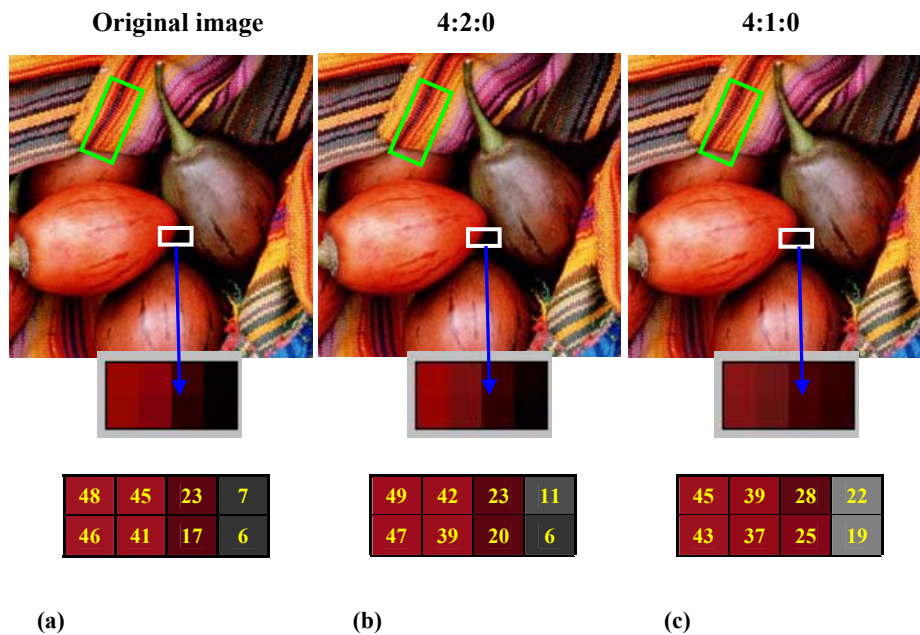
RGB luminance values representing the amount of light. Two colour-difference components, C'_B and C'_R, are formed by subtracting luma component from nonlinear blue and red signals

respectively (i.e. $B' - Y'$ and $R' - Y'$). It was assumed that $C'_B C'_R$ signals produced in this manner should not include luminance information.

If there is no crosstalk between Y' signal and each of C'_B and C'_R signals, there is neither presence of luminance information in $C'_B C'_R$ signals nor presence of chrominance information in Y' signal. The level of crosstalk was checked using an original image seen in Fig. 33(a). $Y' C'_B C'_R$ values for this original image were calculated by means of the encoding scheme given in Fig. 32. Only the two colour-difference signals, $C'_B C'_R$, were subsampled in the format of 4:2:0 and 4:1:0 over an entire image. The subsampled $C'_B C'_R$ signals at 1/4 and 1/8 the original ratio and the Y' signals were then converted back to the RGB values with the original pixel sampling that construct the resultant images shown in Figs. 33(b) and 33(c). A conventional interpolation method was applied to the conversion process back to the original pixel sampling for the subsampled $C'_B C'_R$ signals. The lightness values for the original image shown in Fig. 33(a) and those for the $C'_B C'_R$-subsampled images shown in Figs 33(b) and 33(c) were obtained using CIECAM02-UCS colour appearance model [36]. The area represented by a white box in each of the original and the $C'_B C'_R$-subsampled images is shown in its magnified form (at the bottom of each image) with the calculated lightness values.

FIGURE 33

**The lightness values for the magnified area (indicated by a white box) (a) in the original image
and in the output images that were constructed after subsampling C′B and C′R
in the format of (b) 4:2:0 and (c) 4:1:0**

Original image	4:2:0	4:1:0

48	45	23	7
46	41	17	6

49	42	23	11
47	39	20	6

45	39	28	22
43	37	25	19

(a)   (b)   (c)

Comparing the lightness values for the original image (see Fig. 33(a)) with those for the $C'_B C'_R$-subsampled image in the format of 4:1:0 (see Fig. 33(c)), apparent differences are observed in the right-side four darker pixels. In the $C'_B C'_R$-subsampled image in the format of 4:2:0 (see Fig. 33(b)), slightly increased lightness values are observed in two of the right-side four pixels compared with those in the original image, but the lightness change is not large enough to impact image quality. For the original image seen in Fig. 33(a), a clear boundary of the red fruit can be viewed due to the large lightness-difference between the red fruit and its adjacent dark background. This tendency is not remained for the $C'_B C'_R$-subsampled image seen in Fig. 33(c), leading to a blurred boundary. If the colour-difference signals do not contain luminance information, the lightness values of the original image should not be changed against the variations in the colour-difference values arising

from the subsampling process. However, this is not the case for the $C'_B C'_R$-subsampled images shown in Figs. 33(b) and 33(c). Aforementioned findings demonstrate the level of crosstalk occurring between the luma component and each of the colour-difference components. Inferentially, the existence of crosstalk is caused by applying the luminance coefficients to nonlinear $R'G'B'$ signals for producing the luma component, and by subtracting this luma component from nonlinear B' and R' signals for producing the colour-difference components. The crosstalk problems resulted from these reasons can be propagated if subsampling and compression processes are applied to inaccurate luma and colour-difference signals. The best way to prevent transferring errors originated from the existence of crosstalk is to derive a new encoding scheme that can separate luminance and chrominance information accurately. The current proposal attempted to develop such a new encoding scheme for UHDTV applications.

1.2 A new encoding scheme for UHDTV programme production

1.2.1 Derivation of a matrix producing luma and colour-difference signals

A new encoding scheme was developed using linear colorimetric signals, *XYZ* tristimulus values, which are widely used to specify colours [38]. The reason to choose *XYZ* signals instead of *RGB* signals is for constructing video signals in a perceptual colour space. This can effectively lead to eliminate crosstalk problems that Recommendation ITU-R BT.709 holds. The *XYZ* values are computed from the 1 931 colour matching functions that indicate human visual sensitivities against long-, medium-, and short-wavelength regions in the visible light of 360-830 nm. Figure 34 shows the 1 931 $\bar{x}\,\bar{y}\,\bar{z}$ colour matching functions. The three curves corresponding to *XYZ* are seen to be overlapped one another. Two components (C_{RG} and C_{YB}) having red-green and yellow-blue colour-difference information were generated by separating *X* and *Y*, *X* and *Z*, and *Y* and *Z* signals. One component (*A*) having luminance information was formed from *Y*, because the $\bar{y}$ colour matching function has almost the same shape as photopic spectral luminous efficiency curve $V(\lambda)$ against different wavelengths. The three signals $AC_{YB}C_{RG}$ derived in the present proposal correspond to $Y'C'_B C'_R$ defined in Recommendation ITU-R BT.709.

FIGURE 34

CIE 1931 Colour matching functions
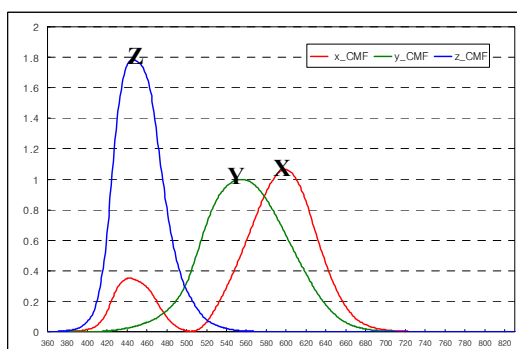
Picture information obtained after capturing an original scene using a video camera can be represented by linear *RGB* values in the range of 0 – 1. The colorimetric meanings for the picture information described using the *RGB* values can be determined according to the reference *RGB* primaries introduced in Table 10 and D65 of the reference white point. The conversion matrix from linear *RGB* values to linear *XYZ* encoding components is shown in equation (1).

TABLE 10

The chromaticity coordinates of the previously proposed RGB-primary set in May 2009 and the modified RGB-primary set in this proposal

Chromaticity coordinates (CIE 1931 xy and CIE 1976 u′v′) of reference primaries	The previously proposed RGB-primary set (May, 2009)		The modified RGB-primary set in this proposal (Nov., 2009)	
	x (u′)	y (v′)	x (u′)	y (v′)
Red	0.7006 (0.5399)	0.2993 (0.5190)	0.7006 (0.5399)	0.2993 (0.5190)
Green	0.0952 (0.0279)	0.9017 (0.5954)	0.1625 (0.0529)	0.8012 (0.5867)
Blue	0.1350 (0.1682)	0.0400 (0.1121)	0.1314 (0.1599)	0.0459 (0.1256)

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.64650 & 0.13483 & 0.16908 \\ 0.27619 & 0.66475 & 0.05906 \\ 0.00009 & 0.03012 & 1.05859 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}
$$
(1)

The linear *XYZ* encoding components are normalized using $X = 0.9504$, $Y = 1$, and $Z = 1.0888$ of the D65 reference white that is described in equation (2).

$$
\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} 1/0.9504 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/1.0888 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}
$$
(2)

The normalized $X_r Y_r Z_r$ encoding components are then transformed onto nonlinear normalized $X_r' Y_r' Z_r'$ encoding components using a nonlinear function introduced in equation (3).

$$
E' = \begin{cases} 4.5E & , 0 \le E < 0.0181 \\ 1.0993E^{0.45} - 0.0993, & 0.0181 \le E \le 1 \end{cases}
$$
(3)

where E is each of normalized $X_r Y_r Z_r$ encodingcomponents, and E' is each of nonlinear normalized $X_r' Y_r' Z_r'$ encoding components.

Luma and colour-difference components ($AC_{YB}C_{RG}$) are finally generated from nonlinear normalized $X_r' Y_r' Z_r'$ encoding components using the matrix given in equation (4) and then are digitized using a series of equations shown in equations (5)-(7).

$$
\begin{bmatrix} A \\ C_{YB} \\ C_{RG} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -0.325 & -0.325 & 0.65 \\ 0.64759 & -0.64719 & -0.0004 \end{bmatrix} \begin{bmatrix} X_r' \\ Y_r' \\ Z_r' \end{bmatrix}
$$
(4)

$$DA = INT\left[(219 \times A + 16) \times 2^{n-8}\right]$$

(5)

$$DC_{YB} = INT\left[\left(224 \times \left(\frac{C_{YB} + 0.53235}{0.97361} - 0.5\right) + 128\right) \times 2^{n-8}\right]$$

(6)

$$DC_{RG} = INT\left[\left(224 \times \left(\frac{C_{RG} + 0.29647}{0.52550} - 0.5\right) + 128\right) \times 2^{n-8}\right]$$

(7)

where n is the number of bits to be represented, and $INT[x]$ gives 0 for $0 \le$ fractional part of $x < 0.5$, and 1 for $0.5 \le$ fractional part of $x < 1$.

The matrix (given in equation (4)) to produce the luma and colour-difference signals was derived so as to satisfy the following three conditions: (1) minimization of correlation in each of AC_{YB}, AC_{RG}, and $C_{YB}C_{RG}$ signal pairs, (2) maintenance of A signal having luminance information regardless of subsampling $C_{YB}C_{RG}$ signals, and (3) for neutral colours, C_{YB} and $C_{RG} = 0$. When A and $C_{YB}C_{RG}$ signals were computed from the matrix of equation (4) reflecting the three conditions of (1) – (3), luma and colour-difference signals having inappreciable crosstalk could be obtained. The yellow-blue colour-difference component C_{YB} was obtained by separating Z_r' from each of X_r' and Y_r'. The red-green colour-difference component C_{RG} was derived by separating X_r' from each of Y_r' and Z_r'. When separating $X_r'Y_r'Z_r'$ encoding components one another, the coefficients multiplied to individual $X_r'Y_r'Z_r'$ encoding components were optimized in order to fulfil the three constraints (1), (2), and (3).

TABLE 11

A colour data set used to derive a matrix computing luma and colour-difference signals

Category	Source	Number of colours	Reference
Real surface colour	Pointer	576	[29]
	SOCS	53 499	[30]
Standard colour space	Recommendation ITU-R BT.709	4 080	[33]
	NTSC	4 080	[35]
	Adobe RGB	4 080	[35]
Flat panel display	LCD	4 080	[31]
	AMOLED	4 080	[31]
Digital cinema	Reference Projector	4 080	[32]
Natural image	Corel image database - 18 categories[1] × 2 images - Pixel resolution for the selected images = 384 (H) ×256 (V)	384 × 256 × 36 = 3 538 944	

[1] The 18 categories are (1) English country gardens, (2) evening skies, (3) food, (4) food and dining, (5) food and drinking, (6) forest, (7) forest floor, (8) forests and trees, (9) garden ornaments and architecture, (10) gardens of Europe, (11) glaciers and mountains, (12) hawks and falcons, (13) interior design, (14) kitchen and bathroom, (15) office interiors, (16) surfside, (17) surfs up, and (18) beach.

In the derivation of the matrix of equation (4), a huge set of colours was used to encompass all the real-world surface colours and a wide range of colour combinations occurring in natural scenes.

Table 11 introduces this data set that has a total number of 3,617,499 colours collected from five categories including wider colour gamut than the conventional TV systems'. For the three categories – "standard colour space", "flat panel display", and "digital cinema", 4 080 colours were sampled among the reproducible colours by the RGB primaries defined in each of Recommendation ITU-R BT.709, Adobe RGB, NTSC, AMOLED, LCD, and digital cinema reference projector. For the category of "natural image", 36 images of diverse contents were selected from Corel image database.

1.2.2 An investigation of crosstalk existing between the new luma and colour-difference channels ($AC_{YB}C_{RG}$) that are produced by the new encoding scheme in this proposal

Correlation coefficients computed from each of three pairs of AC_{YB}, AC_{RG}, and $C_{YB}C_{RG}$ using the data set introduced in Table 11 were all close to zero, i.e. 0.09, 0.15, and 0.07. This proves that the derived matrix given in equation (4) could effectively create independent luma and colour-difference signals, resulting in generation of a nearly orthogonal video-signal set.

An original image shown in Fig. 35(a) was used to assess the level of crosstalk arising from luma and colour-difference signal pairs (AC_{YB} and AC_{RG}). This is the same test-image used to evaluate the level of crosstalk occurring from $Y'C'_B$ and $Y'C'_R$ pairs in Recommendation ITU-R BT.709 in § 1.1. The $AC_{YB}C_{RG}$ values for this original image were calculated from the linear XYZ encoding signals by means of the new encoding scheme described from equation (1) to equation (7). Only the two colour-difference signals, $C_{YB}C_{RG}$, were subsampled in the format of 4:2:0 and 4:1:0 over an entire image. The subsampled $C_{YB}C_{RG}$ signals at 1/4 and 1/8 the original ratio and the A signals were then converted back to the RGB values with the original pixel sampling that construct the resultant images shown in Figs 35(b) and 35(c). A conventional interpolation method was applied to the conversion process back to the original pixel sampling for the subsampled $C_{YB}C_{RG}$ signals. The lightness values for the original image shown in Fig. 35(a) and those for the $C_{YB}C_{RG}$-subsampled images shown in Figs 35(b) and 35(c) were obtained using CIECAM02-UCS colour appearance model [36]. The area represented by a white box in each of the original and the $C_{YB}C_{RG}$-subsampled images is shown in its magnified form (at the bottom of each image) with the calculated lightness values.

The lightness values of the $C_{YB}C_{RG}$-subsampled images shown in Figs 35(b) and 35(c) are almost identical to those of the original image seen in Fig. 35(a) regardless of the subsampling format. As a result, the clear boundary of the red fruit against its dark background and the apparent dark stripe line against its neighbouring orange-yellow background (indicated inside the green box) is remained in the $C_{YB}C_{RG}$-subsampled images compared to the original image. On the contrary, significant lightness discrepancies and varied dark stripe line were found in Fig. 33(c) compared to Fig. 33(a) in § 1.1. The lightness comparison results given in Figs 35(a) to 35(c) indicate that the new encoding scheme can offer the luma and colour-difference signals having inappreciable crosstalk. In other words, it is believed that A signal contains tiny chrominance information, and C_{YB} and C_{RG} signals have inconsiderable luminance information. This fact may enable subsampling ratio in colour-difference channels to be increased with an imperceptible negative effect on image quality. For example, 4:1:0 subsampling format can be applicable to UHDTV systems when the newly developed encoding scheme is used for subsampling colour-difference signals. This is supported by comparing the image quality and the lightness values between the $C_{YB}C_{RG}$ subsampled image in Fig. 35(c) and the original image in Fig. 35(a). Additionally, improved image quality is likely to be obtained when the currently widely used 4:2:0 format is applied to the $C_{YB}C_{RG}$ signals that are created by the new encoding scheme suggested in this work.

FIGURE 35

**The lightness values for the magnified area (indicated by a white box) (a) in the original image
and in the output images that were constructed after subsampling C$_{YB}$ and C$_{RG}$ signals
in the format of (b) 4:2:0 and (c) 4:1:0**
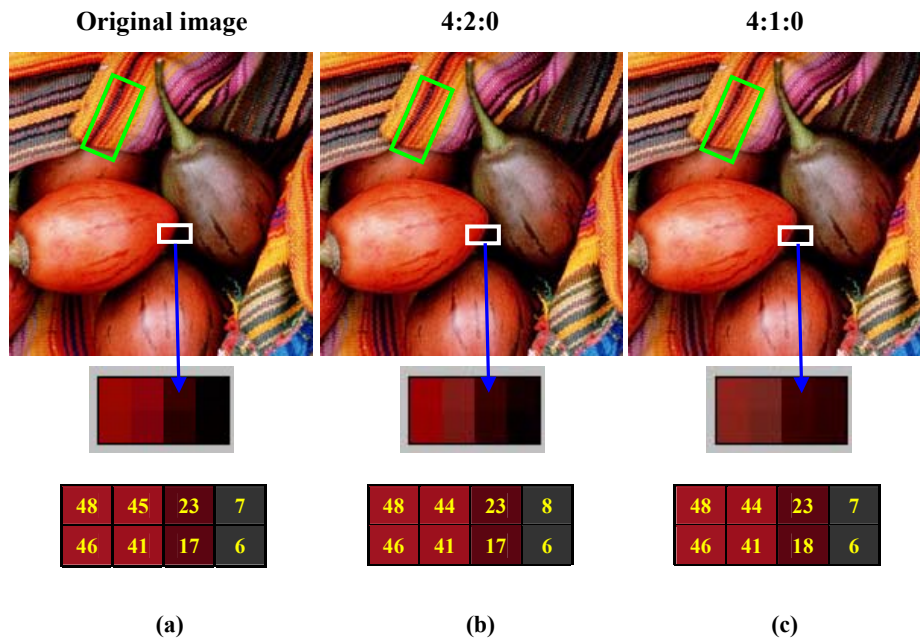
2 Conclusions

This Appendix provided scientific foundation applied to propose the new RGB-primary set and the new colour encoding-scheme that are believed to be suitable for UHDTV programme production. The new RGB-primary set was derived in favour of inclusion of real-world surface colours and in consideration of the characteristics of AMOLED and LCD using which UHDTV programmes could be displayed. The new encoding scheme was developed from *XYZ* tristimulus values reflecting human visual properties and aimed at producing luma and colour-difference signals that were as much as possibly orthogonal. It was proved that luminance and chrominance information could be almost accurately separated one another by the new encoding scheme. This result can increase chrominance subsampling ratio (e.g. 4:1:0) without an adverse effect on image quality, or enhance image quality when the widely accepted subsampling-format of 4:2:0 or 4:2:2 are used. The influence of the newly developed luma and colour-difference signals on image data redundancy will be further investigated so that the new encoding scheme will be able to be verified for a video-signal processing chain.

References

[29]   M. R. Pointer, "The gamut of real surface colors", Col. Res. Appl., 5(3), p.145 (1980)

[30]   ISO/TR 16066, "Graphic technology – Standard object color spectra database for colour reproduction evaluation" (2003)

[31]   Samsung Electronics Products

[32]   Digital Cinema System Specification, ver. 1.0, Digital Cinema Initiatives, LLC Member Representatives Committee (2005)

[33]   Recommendation ITU-R BT.709-5, "Parameter values for the HDTV standards for production and international programme exchange", ITU, Geneva, Switzerland (2002)

[34]   Adobe RGB® Color Image Encoding, Adobe Systems Incorporated (2005)

[35]   SMPTE 170M-2004, "Composite analog video signal – NTSC for studio applications" (2004)

[36]   M.R. Luo, G. Cui and C. Li, "Uniform color spaces based on CIECAM02 color appearance model", Col. Res. Appl., 31(4), p.320 (2006)

[37]   CIE Publication 15:2004, Colorimetry, Central bureau of the CIE, Vienna, Austria (2004)

[38]   R.W.G. Hunt, Measuring Color, Fountain Press, Kingston-upon-Thames, England (1998).

ATTACHMENT 4

A study on chroma sub-sample and signal equations

1   Coded signal

[Input, 19] and [Input, 23] call for the inclusion of only Y, C_{YB}, C_{RG} for coded signals in its Table 5. It is ambiguous whether the RGB 4:4:4 format is included because they do not specifically mention the RGB 4:4:4 format. As [Input, 25] clarifies this point and includes the RGB 4:4:4 format, it is believed that both RGB and YCC (luminance and colour-difference) formats should be specified for the coded signal.

2   Chroma sub-sample ratio

[Input, 19] proposes 4:4:4, 4:2:2, 4:2:0, and 4:1:0 formats. [Input, 23] proposes to exclude 4:2:0 and 4:1:0. It is reasonable to exclude 4:1:0 from the production formats [39]. However, regarding 4:2:0, it should be noted that 4:2:0 forms a square sampling lattice for colour difference signals for progressive scanning, as 4:2:2 does for interlaced scanning, and that 4:2:0 helps to reduce the data rate while keeping the picture quality high even though it adds a number of circuits and incurs a delay of several lines. It is thus proposed to include it along with the 4:4:4 and 4:2:2 formats.

3   Constant luminance issue

[Input, 19] proposes new equations for luminance and colour difference signals. [Input, 23] provides supportive evidence for the equations by comparing them with the equations in Recommendation ITU-R BT.709. The intention of the new equations is to eliminate the "crosstalk" problem between the luminance and colour difference signals. A major aspect of this problem is the well-known "non-constant luminance problem". The non-constant luminance problem is due to the derivation of the luminance signal from gamma-corrected R, G, B signals. Therefore, the determination of signal equations resolves itself into determining the equation for luminance and determining the equation for colour differences.

The luminance equation proposed in [Input, 19] and supported by [Input, 23] solves the non-constant luminance problem by deriving the luminance signal from linear RGB signals and applying gamma correction to the luminance signal. The new luminance equation is supposed to have the desired effect upon the non-constant luminance problem regardless of which equation is chosen for colour difference signals.

4   Colour difference equations

There are two options for colour difference equations; one is the conventional simple colour differences between blue and luminance (B' − Y') and between red and luminance (R' − Y'), and the other is proposed in [Input, 19], called C_{YB} and C_{RG}. Depending on whether the condition is of non-constant luminance or constant luminance, we have the following three options to consider.

A   Rec. ITU-R BT.709 (Non-constant luminance, B'-Y' and R'-Y')

$$Y' = rR' + gG' + bB'$$

$$C_B{}' = a_b(B' - Y')$$

$$C_R{}' = a_r(R' - Y')$$

B   [Input, 19] (Constant luminance, C_{YB} and C_{RG})

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.64650 & 0.13483 & 0.16908 \\ 0.27619 & 0.66475 & 0.05906 \\ 0.00009 & 0.03012 & 1.05859 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} 1/0.9504 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/1.0888 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\begin{bmatrix} A \\ C_{YB} \\ C_{RG} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -0.325 & -0.325 & 0.65 \\ 0.64759 & -0.64719 & -0.0004 \end{bmatrix} \begin{bmatrix} X'_r \\ Y'_r \\ Z'_r \end{bmatrix}$$

C   Alternative scheme (Constant luminance, B'-Y' and R'-Y')

$$Y' = (rR + gG + bB)'$$

$$C_B{}' = a_b(B' - Y')$$

$$C_R{}' = a_r(R' - Y')$$

The constant luminance option which is applied to equations B and C has a potential problem. As shown in Fig. 36, the positive and negative excursions from the achromatic level differ. The following two approaches were considered to cope with this problem. Other ways, such as an approach that allocates the middle code to the achromatic signal while different gains are used for positive and negative excursions, are out of consideration due to their complexity.

Approach 1: The middle code is allocated to the achromatic signal. The gain is based on the largest excursion.

Approach 2: The gain is based on the peak-to-peak excursion. The code for the achromatic signal is biased from the middle.

Although [Input, 19] takes approach 2, our experiments on codecs complying with MPEG-2 Video and MPEG-4 AVC showed this approach caused picture quality degradation. Approach 1 would be desirable since it would enable the current practice to continue and could be harmonized with other applications if the decrease in efficiency of digital code usage is marginal. Table 12 lists the colour errors of approach 1 and 2 in case of equation C with current Rec. ITU-R BT.709. Although the colour errors of approach 1 are not the least among the three, they are quite acceptable.
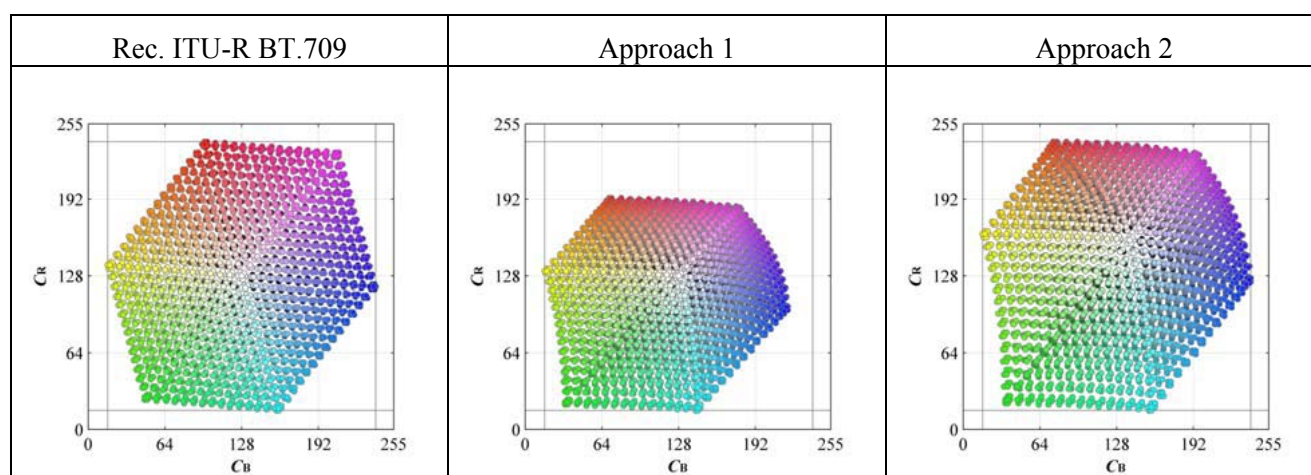
FIGURE 36

Code usage

| Rec. ITU-R BT.709	Approach 1	Approach 2

TABLE 12

Colour error expressed in ΔE_{ab^*}

10-bit	Rec. BT.709	Approach 1	Approach 2
Average	0.14554	0.17678	0.15569
Maximum	0.89911	1.1716	1.1416

These three sets of colour difference equations have been evaluated by using the metric of root mean square error (RMSE) and the images used in [Input, 23]. RMSEs are calculated for the lightness of the image for the cascade of one through four. The results are shown in Fig. 37. In the calculation, approach 1 is used for the code assignment discussed in § 3.3. The RMSE for equations B and C are quite small and almost the same, whereas equation A suffers from the non-constant luminance problem reported in [Input, 23]. This means that the improvement of the RMSE is mostly due to the constant-luminance scheme, not due to the selection of colour difference equations. In other words, the non-constant luminance problem can only be solved by applying a gamma correction for the luminance signal after deriving it from linear R, G, and B signals.
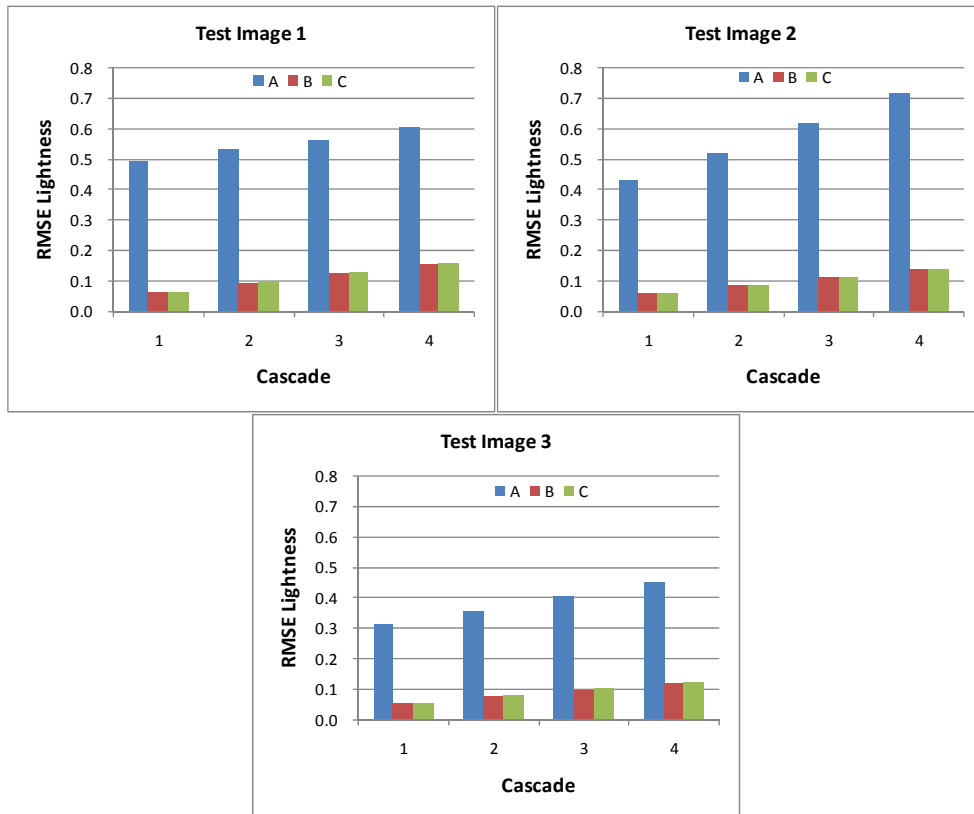
This means there is little advantage in using new colour difference equations, i.e., equation B, which departs from current practice. Equation C can solve the non-constant luminance problem, and it does not deviate much from current practice.

FIGURE 37

RMSE of lightness

(a) 4:2:2
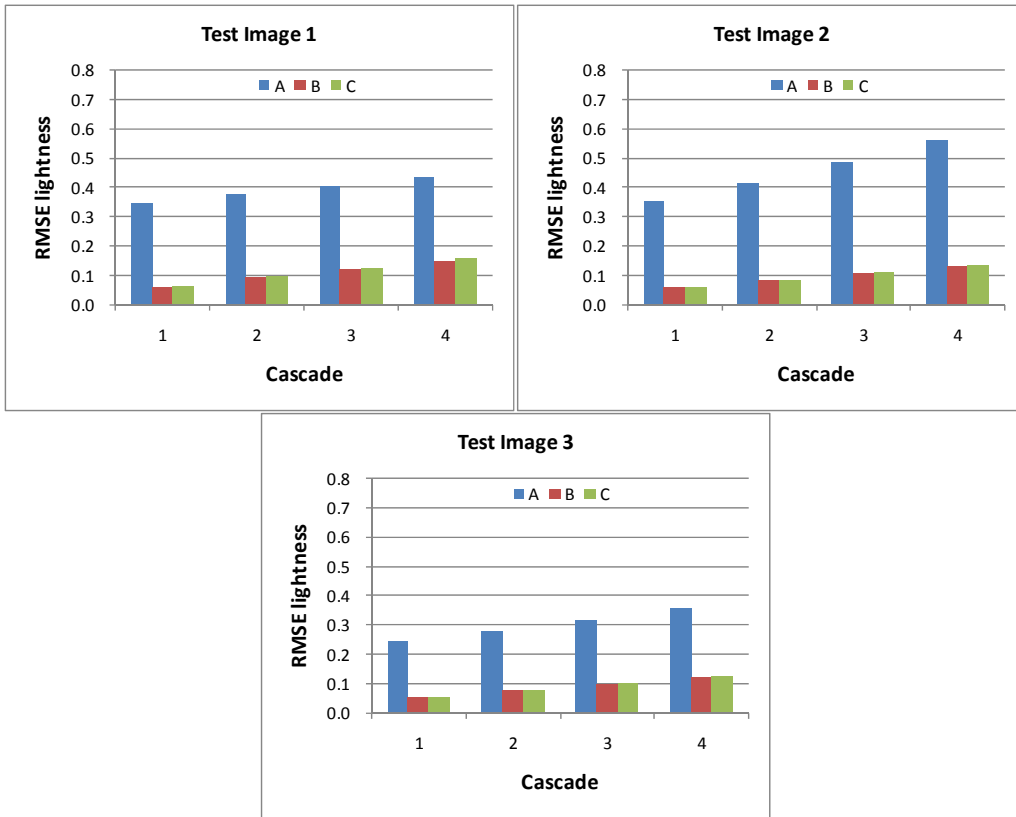
(b) 4:2:0

5 Conclusion

The following proposal is based on the above discussion.

–   Chroma sub-sample ratio:

  4:4:4 (including R, G, B), 4:2:2, and 4:2:0

–   Signal equations:

$$Y' = (0.2627\ R + 0.6780\ G + 0.0593\ B)'$$

$$C_B' = \frac{B' - Y'}{1.9403}$$

$$C_R' = \frac{R' - Y'}{1.7182}$$

References

[39]   S. Baron and D. Wood, "The Foundation of Digital Television: The Origins of the 4:2:2 Component Digital Television Standard," SMPTE Motion Imaging Journal, pp. 327-334, September 2005.

ATTACHMENT 5

Evaluation results for the compromise signal format proposed in [Input, 41] compared with the signal format of Recommendation ITU-R BT.709

Table 13

The compromise [Input, 41] between the two signal formats described in Attachment 1

Parameter	Values	
Signal format	$R'G'B'$	Luma (Y') and colour difference (C_B' and C_R')
Derivation of luma signal Y'	$$Y' = \left(0.2627R + 0.6780G + 0.0593B\right)'$$ where linear Y signal is converted to nonlinear Y' signal using the nonlinear function introduced in the following row.	
Nonlinear transfer function	$$E' = \begin{cases} 4.5E, & 0 \le E < \beta \\ \alpha E^{0.45} - (\alpha-1), & \beta \le E \le 1 \end{cases}$$ where E is voltage normalized by the reference white level and proportional to the implicit light intensity that would be detected with a reference camera colour channel R, G, B; E' is the resulting nonlinear signal. α = 1.099 and β = 0.018 for 10-bit system α = 1.0993 and β = 0.0181 for 12-bit system	
Derivation of colour difference signals C_B' and C_R'	$$C_B' = \begin{cases} \dfrac{B'-Y'}{1.9404}, & -0.9702 \le B'-Y' \le 0 \\ \dfrac{B'-Y'}{1.5816}, & 0 < B'-Y' \le 0.7908 \end{cases}$$ $$C_R' = \begin{cases} \dfrac{R'-Y'}{1.7184}, & -0.8592 \le R'-Y' \le 0 \\ \dfrac{R'-Y'}{0.9936}, & 0 < R'-Y' \le 0.4968 \end{cases}$$	

Bit rate reduction effects

The influence of different encoding methods on the quality of compressed images and compression efficiency was investigated using two moving images (1920H ×1080V) shown in Figs 38(a) and 38(b) that were purchased from The Institute of Image Information and Television Engineers in Japan (http://www.ite.or.jp/data/c/?p=search&q=HDTV%20test%20materials). Actually, NHK Engineering Co. produced these moving images. The test image seen in Fig. 38(a) includes many colourful objects while that seen in Fig. 38(b) appears to be achromatic, i.e., much less colourful scene. These two sets of test sequence were provided in TIFF format composed of nonlinear Recommendation ITU-R BT.709 $R'G'B'$ for 59.94 Hz/interlace scanning. Only 61 frames in the original test sequences were used to analyse compression coding efficiency and the quality of the compressed images. The $R'G'B'$ data in the original images were converted to the conventional video signals of the Recommendation ITU-R BT.709 and the new video signals of the DNR ITU-R BT.[IMAGE-UHDTV]. After subsampling and low-pass filtering colour-difference signals in the

format of 4:2:2, the chroma-subsampled images were then compressed using the reference software JM12.0 [40-43]. The encoder configurations applied to compress the test sequences are as follows.

1)   High profile

2)   GOP structure: I-B-B-P-B-B-P

3)   Entropy coding mode: CABAC

4)   R-D optimization enabled

5)   Number of reference and search range: 1 and 64

6)   QP for I/P/B slice = 22/23/24, 27/28/29, 32/33/34, 37/38/39

FIGURE 38

Two moving test images

Test Sequence 1   **Test Sequence 2**

(a)   (b)

Different luma and colour-difference signals, which were produced according to the two different signal formats defined in the Recommendation ITU-R BT.709 and the DNR ITU-R BT.[IMAGE-UHDTV], were used on the compression process. Instead of PSNR-luma, the PSNR-lightness L^*, -chroma C^*, -hue h and -colour difference ΔE^*_{ab} were therefore used to measure objective quality differences arising between the original and its compressed images. Both luma and colour-difference components are manipulated on the compression process whereas only colour-difference signals are altered on the chroma subsampling process. This fact indicates that both luminance and chrominance information in the original would be changed after the compression process. Thus, variations in the chroma and hue domains as well as those in the lightness domain were calculated between the original and its compressed versions.

Figs 39(a) and 39(b) show the changes in the PSNR-lightness, -chroma, -hue and -colour difference values against bit-rates (kbits/sec) at 59.94 Hz for the two test sequences seen in Figs 38(a) and 38(b). The BD-PSNR and BD-rate values were calculated between the data points of the Recommendation ITU-R BT.709 and those of the DNR ITU-R BT.[IMAGE-UHDTV] plotted in Figs. 39(a) and 39(b). And the calculated results are introduced in Table 14 [44]. The mean of the BD-PSNR and BD-rate values were calculated against the very colourful test sequence 1 and the neutral test sequence 2 and are also described in Table 14. The effects in the change of video signals produced by different encoding methods on the quality of compressed images can be summarised in the following.

1)   The DNR ITU-R BT.[IMAGE-UHDTV] in Table 14 shows negative mean BD-rate
          (about -8 – -10 %) and positive mean BD-PSNR (about 0.2 – 0.3 dB) values in the
          chroma, hue and colour difference domains. In Figs 39(a) and 39(b), the green lines
          indicating the DNR method are located at higher vertical places in the PSNR-chroma,
          hue and colour-difference domains. These results suggest that approximately 8 – 10 %
          less bit-rate is required for the DNR method than for the Recommendation ITU-R
          BT.709 supposing that the same PSNR values are achieved between the two methods
          proposed in the DNR and the Recommendation ITU-R BT.709.

2)   Both BD-rate and BD-PSNR values in the lightness dimension are close to zero: 0.6 %
          of the mean BD-rate and 0 dB of the mean BD-PSNR.

In conclusion, applying the new video signals of the DNR to the H.264 codec can save 8 – 10 %
bit-rate compared with the Recommendation ITU-R BT.709 on the condition that similar
image-quality after the compression stage is perceived.

TABLE 14

**The calculated results for the BD-PSNR and BD-rate measures that were obtained from the colour encoding
methods defined in the DNR ITU-R BT.[IMAGE-UHDTV] in comparison with that
of the Recommendation ITU-R BT.709**

	Lightness		Chroma		Hue		Colour Difference	
	BD-rate	BD-PSNR	BD-rate	BD-PSNR	BD-rate	BD-PSNR	BD-rate	BD-PSNR
Test sequence 1 (very colourful appearance)	-0.4	0.0	-7.2	0.2	-8.3	0.2	-7.0	0.2
Test sequence 2 (neutral appearance)	1.5	-0.1	-12.4	0.3	-9.6	0.1	-9.2	0.2
Mean of two test sequences	**0.6**	**0.0**	**-9.8**	**0.3**	**-9.0**	**0.2**	**-8.1**	**0.2**

FIGURE 39

The PSNR-lightness, -chroma, -hue and -colour difference values against the bit-rates obtained after compressing4:2:2 subsampled images for (a) Test Sequence 1 and (b) Test Sequence 2
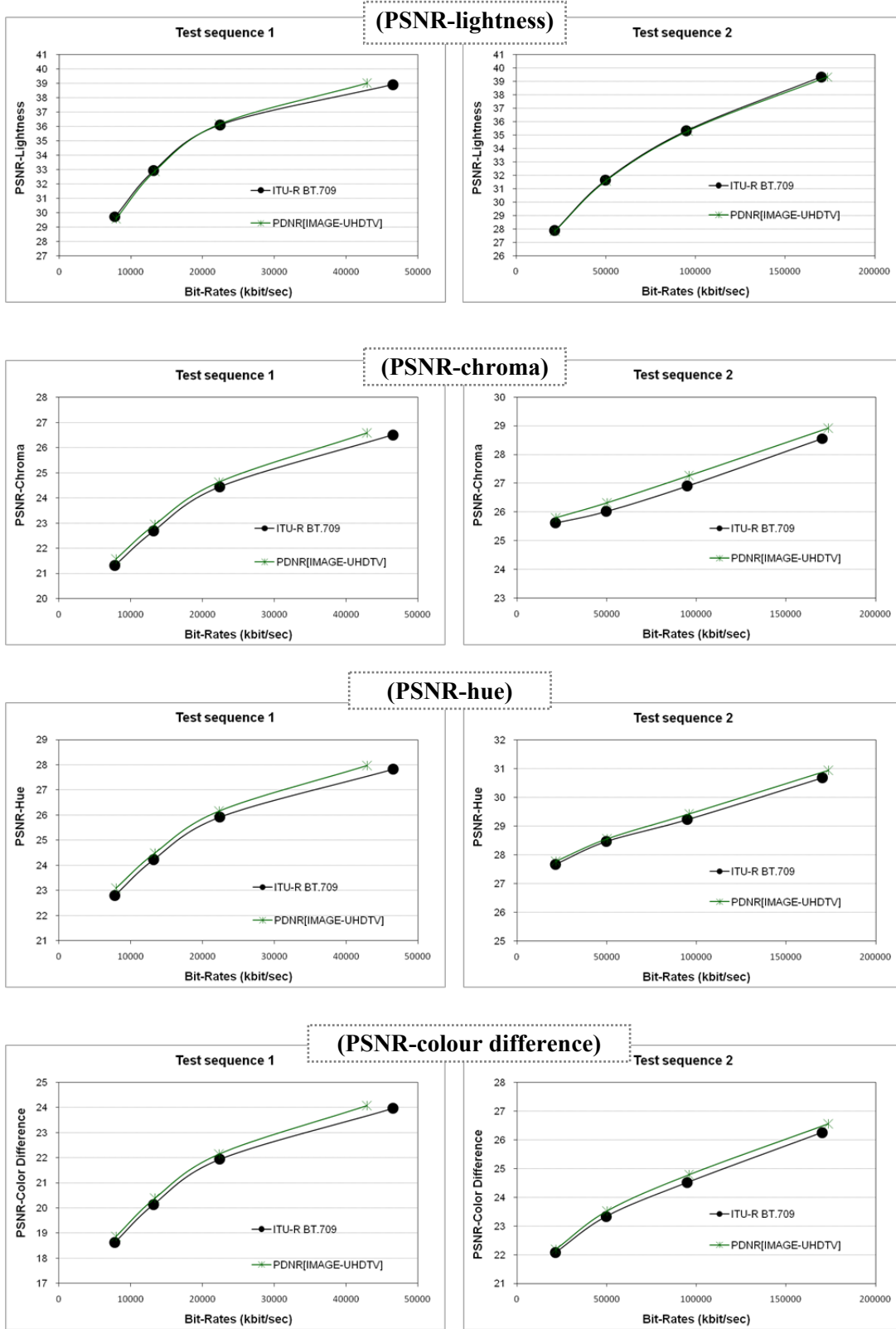
(a)   (b)

Analysis of colour changes owing to subsampling chrominance signals

If any colour stimuli are mixed (subsampled), the colour appearance on the resultant stimulus is perceived differently from that on each of the original stimuli to be mixed (subsampled). There is not a video signal set that can preclude this colour appearance change. Therefore, it will be the most appropriate to choose the video signal set providing fewer occasions where significant colour changes from the original due to the subsampling process are perceived.

The colour changes from the original after the subsampling process were estimated using 7750 colour mixing cases that were generated from all the possible combinations of 124 colours covering the entire colour gamut defined in the DNR ITU-R BT.[IMAGE-UHDTV]. Figure 40 shows the chosen 124 colours in the CIE u'v' chromaticity diagram. The different colours to be subsampled in each 7750 stimuli were alternately placed on the vertical pixel basis, i.e., the highest frequency. Some of the test stimuli are illustrated in Fig. 41. The original colours were down-sampled in the format of 4:2:0. Then, the down-sampled colours were reconstructed back to the original resolution. The colour change was measured in terms of PSNR-lightness, chroma and hue that were computed between the original and its reconstructed stimulus on the CIELAB L^*C^*h domain. Different luma and colour-difference signals, which were produced according to the two different signal formats defined in the Recommendation ITU-R BT.709 and the DNR ITU-R BT.[IMAGE-UHDTV], were used for the subsampling process.

FIGURE 40

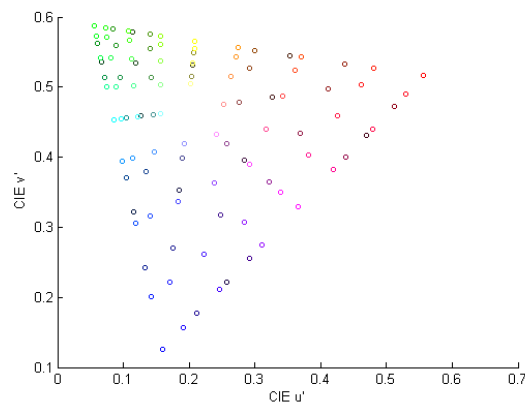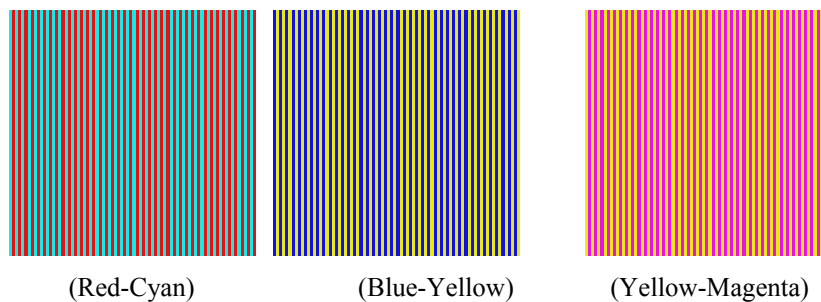**The 124 colours chosen to produce 7750 colour mixing cases**

FIGURE 41

Some colour stimuli used to evaluate colour changes from the original after the subsampling process

(Red-Cyan)   (Blue-Yellow)   (Yellow-Magenta)

The number of colour samples showing less than 40 dB in each of the PSNR-lightness, chroma and hue values was counted. These counted numbers are compared in Table 15 for the two different signal formats of Recommendation ITU-R BT.709 and the DNR ITU-R BT.[IMAGE-UHDTV]. As the counted number is smaller, fewer cases where the significant colour changes from the original are expected. The comparison result in Table 15 clearly indicates that the original luminance information can be preserved more effectively for the signal format of the DNR (811) than for that of Recommendation ITU-R BT.709 (6745) after the subsampling process. Similar colour changes in the aspect of chroma and hue properties are however found for both signal formats. In conclusion, the new signal format proposed in the DNR perform better in the maintenance of the original luminance information than the conventional signal format, leading to more preservation of the original detail and edge information and further less image quality degradation after the subsampling process.

TABLE 15

The analysis results of colour changes occurring after the subsampling process using 7750 colour mixing cases

Signal format	PSNR-lightness	PSNR-chroma	PSNR-hue
Recommendation ITU-R BT.709	**6745**	7646	7361
DNR ITU-R BT.[IMAGE-UHDTV]	**811**	7702	7450

References

[40]   Recommendation ITU-T H.264 and ISO/IEC 14496-10, Advanced Video Coding for Generic Audiovisual Services, May 2003.

[41]   Recommended simulation common conditions for coding efficiency experiments, Document VCEG-AA10, ITU-T Study Group 16 Question 6, VCEG, ITU-T, October 2005.

[42]   Color format down-conversion for test sequence generation, MPEG2003/N6295, ISO/IEC JTC1/SC29/WG11, Waikoloa, December 2003.

[43]   Color format up-conversion for video display, MPEG2003/N6296, ISO/IEC JTC1/SC29/WG11, Waikoloa, December 2003.

[44]   An excel add-in for computing Bjontegaard metric and its evolution, Document VCEG-AE-07, ITU-T Study Group 16 Question 6, VCEG, ITU-T, January 2007.

ATTACHMENT 6

Parameter values for UHDTV systems for production and international programme exchange

1 Discussion of the luminance and colour signal formats described in the PDNR

1.1 Current specification

[Input, 41] provides specifications of luminance and colour-difference signals, as listed in Table 16.

TABLE 16

Specifications of luminance and colour-difference signal format in PDNR

Derivation of luma signal Y'	$$Y' = (0.2627\,R + 0.6780\,G + 0.0593\,B)'$$ where linear Y signal is converted to nonlinear Y' signal using the nonlinear function introduced in the following row.
Nonlinear transfer function	$$E' = \begin{cases} 4.5E, & 0 \le E < \beta \\ \alpha E^{0.45} - (\alpha - 1), & \beta \le E \le 1 \end{cases}$$ where E is voltage normalized by the reference white level and proportional to the implicit light intensity that would be detected with a reference camera colour channel R, G, B; E' is the resulting nonlinear signal. $\acute{a} = 1.099$ and $\hat{a} = 0.018$ for 10-bit system $\acute{a} = 1.0993$ and $\hat{a} = 0.0181$ for 12-bit system
Derivation of colour difference signals C'_B and C'_R	$$C'_B = \begin{cases} \dfrac{B' - Y'}{1.9404}, & -0.9702 \le B' - Y' \le 0 \\ \dfrac{B' - Y'}{1.5816}, & 0 < B' - Y' \le 0.7908 \end{cases}$$ $$C'_R = \begin{cases} \dfrac{R' - Y'}{1.7184}, & -0.8592 \le R' - Y' \le 0 \\ \dfrac{R' - Y'}{0.9936}, & 0 < R' - Y' \le 0.4968 \end{cases}$$
Sampling lattice − R', G', B', Y'	Orthogonal, line and picture repetitive co-sited

Sampling lattice − C'_B, C'_R	Orthogonal, line and picture repetitive co-sited with each other. The first (top-left) sample is co-sited with the first Y' samples.		
	4:4:4 system	4:2:2 system	4:2:0 system
	Each has the same number of horizontal samples as the Y' component.	Horizontally subsampled by a factor of two with respect to the Y' component.	Horizontally and vertically subsampled by a factor of two with respect to the Y' component.

1.2 Discussion

Japan concurs that the constant luminance transmission that derives the luminance signal *Y* from 'linear' *RGB* signals is ideal. With respect to the specification of the colour-difference signals, Japan has the following observations.

1.2.1 Low-pass filtering for sub-sampling

It is understood that the colour-difference signals C'_B and C'_R are low-pass filtered and sub-sampled for the 4:2:2 and 4:2:0 systems. It should be noted that since C'_B and C'_R signals are not represented by the linear sum of *R'G'B'* nor *RGB*, C'_B and C'_R signals contain non-linear products. Low-pass filtering of such signals results in unexpected tint, as shown in Fig. 42. The original image shown in Fig. 42 comprises sinusoidal patterns alternating the chroma of red and cyan with seven spatial frequencies of 1, 2/3, 1/2, 1/4, 1/8, 1/12, and 1/16 times the Nyquist frequency and four modulation depths of 1, 3/4, 1/2, and 1/4. Ideally, the sinusoidal patterns with spatial frequencies of 1, 2/3, and 1/2 times the Nyquist frequency are reproduced in grey when C'_B and C'_R are 2:1 sub-sampled.

To avoid the problem of unexpected tint, it is essential to apply low-pass filtering to linear *YBR* signals. The sub-sampled *YBR* signals are then nonlinearly transformed to yield *Y'B'R'* signals. Finally, the colour-difference signals $B'-Y'$ and $R'-Y'$ are produced. An example of the modified process is shown in Fig. 43. The underlined symbols in the processes represent the sub-sampled signals.

FIGURE 42

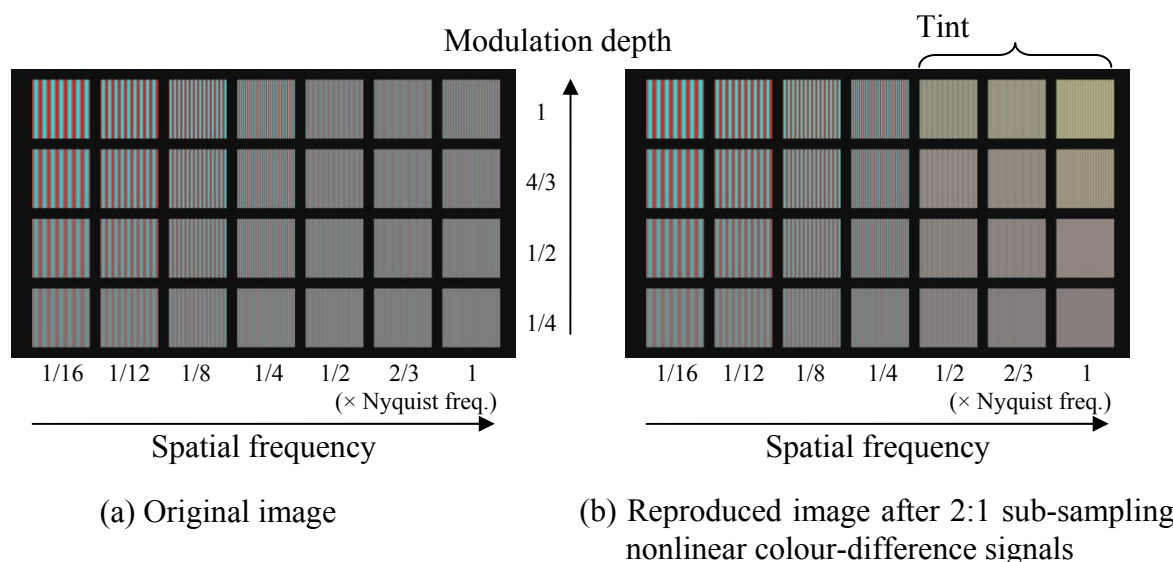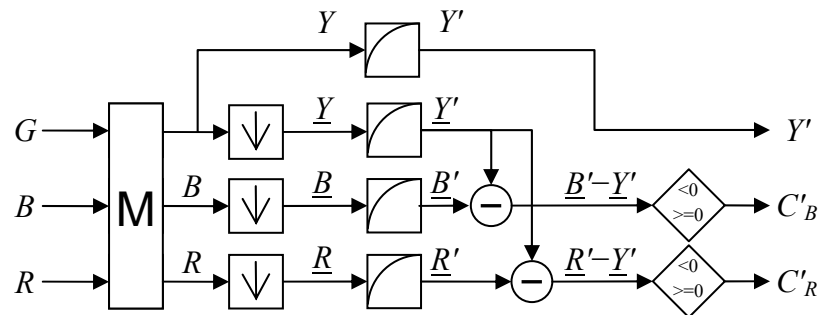**Example of unexpected tint due to low-pass filtering of nonlinear colour-difference signals**

(a) Original image

(b) Reproduced image after 2:1 sub-sampling nonlinear colour-difference signals

FIGURE 43

Example of encoding process to apply sub-sampling to linear signals


1.2.2 Video signal processing using Y', C'_B, and C'_R signals

It is a common practice in SDTV and HDTV environments for video signal processing to be carried out using $Y'C'_BC'_R$ signals. This does not reproduce ideally correct luminance and colour information but results in picture quality that is practically sufficient.

In the case of constant luminance transmission, because of the same reason as that mentioned in § 1.2.1, video processing using $Y'C'_BC'_R$ signals yields slightly worse reproduction of luminance and chrominance information than that produced by the non-constant luminance transmission. Since constant luminance transmission is to improve reproduction of luminance and colour information, this suggests that video processing should be carried out on linear signals such as *YBR* or *RGB* signals that have been converted from $Y'C'_BC'_R$ signals for constant luminance transmission. An example decoding process for $Y'C'_BC'_R$ signals to obtain linear signals is shown in Fig. 44. Either *YBR* or *RGB* may be used as linear signals.

FIGURE 44

Example of decoding process for the constant luminance $Y'C'_BC'_R$ signals
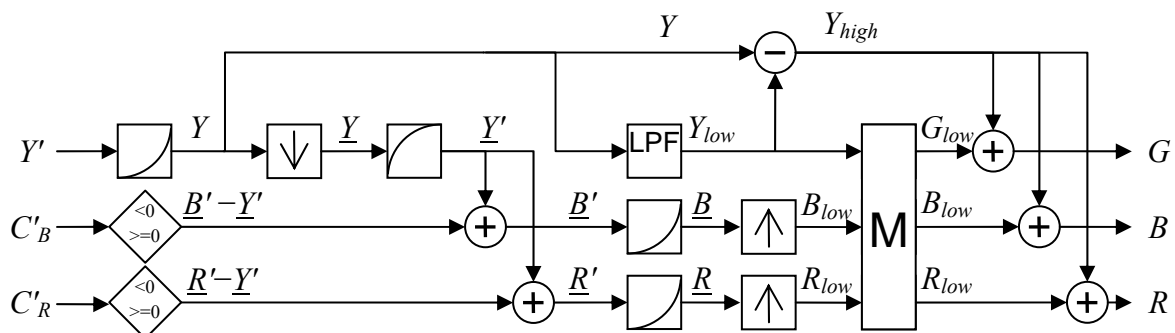
1.2.3 Guidance on encoding/decoding

Although guidance on the encoding/decoding process for $Y'C'_BC'_R$ has not been specified for SDTV and HDTV that employ non-constant transmission, it is advisable to provide guidance on how to encode/decode signals including characteristics of sub-sampling/up-sampling filters and nonlinear transformation to ensure correct colour reproduction for constant luminance transmission. This may be a future work after completing a Recommendation for the studio standard.

2 Alternative luminance and chrominance signal format for constant-luminance transmission

2.1 Proposed set of luminance and colour signals

As discussed in § 1, the current specification described in the PDNR requires improvement. Furthermore, extensive additional circuitry will always be needed for the improvement. To lighten the burden, Japan proposes using the alternative set of signal formats *Y'B'R'*, among which *B* and *R* are sub-sampled for 4:2:2 and 4:2:0 systems. An example process for encoding/decoding this set of signal formats is shown in Fig. 45. The encoded non-linear signals *Y'B'R'* must be linearized for video signal processing.

FIGURE 45

Example of encoding/decoding process for the constant luminance *Y'B'R'* signals

(a) Encoding (b) Decoding
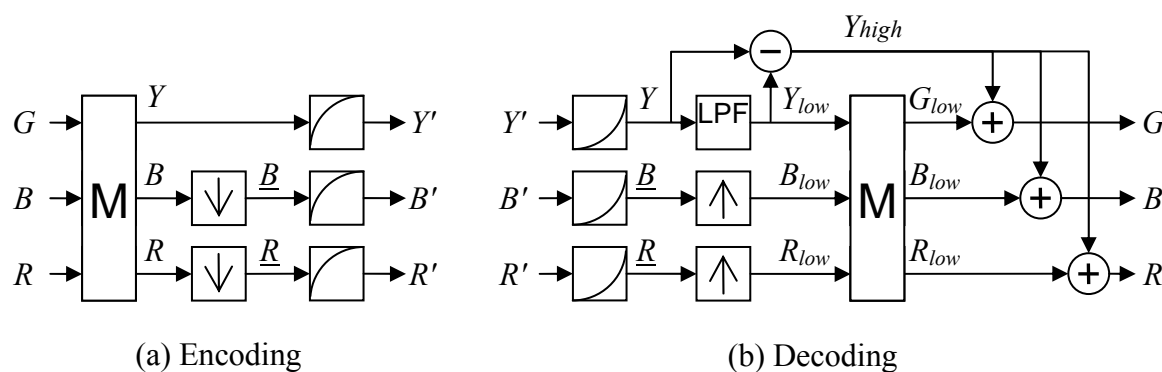
The proposed signal formats are advantageous in terms of the following aspects:

– Required circuitry for encoding/decoding as well as for conversion to/from linear signals is small.

– Bit precision of 1-bit is improved since it does not take a form of colour-difference.

2.2 Consideration of colour-difference signals

Colour-difference signals have been used for studio signals with SDTV and HDTV only because of historical reasons in the current digital environment. Although colour-difference signals may be advantageous in some circumstances, e.g. for compression coding, the use of colour-difference signals is not essential for studio signals that traverse between equipment and between studios using digital interfaces. Even if a different signal format is required for a specific application, it is possible to convert the *Y'B'R'* signals into any other signal format that best suits the application requirements.

3   Alternative luminance and chrominance signal format for non-constant-luminance transmission

Non-constant luminance transmission adopted for SDTV and HDTV has been used without practical problems, although it is not ideal. Moreover, the picture quality produced by video signal processing using nonlinear $Y'C'_BC'_R$ signals is practically sufficient.

It is proposed that the use of a set of signal formats $Y'C'_BC'_R$ for non-constant luminance transmission. It should be noted that this set of signal formats has been included in SMPTE 2036-1.

4   Conclusion

It is proposed that $Y'B'R'$ should be adopted, instead of $Y'C'_BC'_R$, for the signal format for constant-luminance transmission. It is also proposed that $Y'C'_BC'_R$ be adopted as the signal format for non-constant luminance transmission. Proposed modifications to the PDNR are shown in Annex 1. It is also proposed that the PDNR with the proposed modifications be turned into a draft new Recommendation for adoption by Study Group 6.

ANNEX

Proposed modifications to Preliminary Draft New Recommendation ITU-R BT.[IMAGE-UHDTV] FOR A DRAFT NEW RECOMMENDATION

Parameter values for UHDTV systems for production and international programme exchange

Tables 4 and 5 should be modified as follows:

TABLE 4

Signal format

Parameter	Values		
Signal format	$R'G'B'$	Constant luminance $Y'B'R'$	Non-constant luminance $Y'C'_B C'_R$
Derivation of constant luminance signal Y'	$Y' = (0.2627 R + 0.6780 G + 0.0593 B)'$ where linear Y signal is converted to nonlinear Y' signal using the nonlinear transfer function.		
Derivation of non-constant luminance signal Y'	$Y' = 0.2627R' + 0.6780G' + 0.0593B'$		
Nonlinear transfer function	$$E' = \begin{cases} 4.5E, & 0 \le E < \beta \\ \alpha E^{0.45} - (\alpha - 1), & \beta \le E \le 1 \end{cases}$$ where E is the voltage normalized by the reference white level and is proportional to the implicit light intensity that would be detected with a reference camera colour channel R, G, B; E' is the resulting nonlinear signal. $á$ = 1.099 and $â$ = 0.018 for 10-bit system $á$ = 1.0993 and $â$ = 0.0181 for 12-bit system		
Derivation of colour difference signals C'_B and C'_R for non-constant luminance transmission	$$C'_B = \frac{B' - Y'}{1.8814}$$ $$C'_R = \frac{R' - Y'}{1.4746}$$ where Y' is the non-constant luminance signal.		

TABLE 5

Digital representation

Parameters	Values		
Filter	Assumed to have been appropriately filtered to reduce or prevent aliasing upon sampling.		
Coded signal	(R', G', B'), (Y', B', R') *or* (Y', C'_B, C'_R)		
Sampling lattice – R', G', B' for 4:4:4 – Y' for $Y'B'R'$ or $Y'C'_BC'_R$	Orthogonal, line and picture repetitive co-sited		
Sampling lattice – B', R' for $Y'B'R'$ – C'_B, C'_R	Orthogonal, line and picture repetitive co-sited with each other. The first (top-left) sample is co-sited with the first Y' samples.		
	4:4:4 system	4:2:2 system	4:2:0 system
	Each has the same number of horizontal samples as the Y' component.	Horizontally sub-sampled by a factor of two with respect to the Y' component.	Horizontally and vertically sub-sampled by a factor of two with respect to the Y' component.
Coding format	10 or 12 bits per component		
Quantization of R', G', B', Y', C'_B, C'_R	$$DR' = INT\left[(219 \times R' + 16) \times 2^{n-8}\right]$$ $$DG' = INT\left[(219 \times G' + 16) \times 2^{n-8}\right]$$ $$DB' = INT\left[(219 \times B' + 16) \times 2^{n-8}\right]$$ $$DY' = INT\left[(219 \times Y' + 16) \times 2^{n-8}\right]$$ $$DC'_B = INT\left[(224 \times C'_B + 128) \times 2^{n-8}\right]$$ $$DC'_R = INT\left[(224 \times C'_R + 128) \times 2^{n-8}\right]$$		
Quantization levels	10-bit coding		12-bit coding
– Black level DR', DG', DB', DY'	64		256
– Achromatic DC'_B, DC'_R	512		2048
– Nominal Peak DR', DG', DB', DY'	940		3760
DC'_B, DC'_R	64 and 960		256 and 3840
Quantization level assignment	10-bit coding		12-bit coding
– Video data – Timing reference	4 through 1019 0-3 and 1020-1023		16 through 4079 0-15 and 4080-4095

––––––––––––––––