THE SUPER HI-VISION CODEC

Shinichi Sakaida, *Nao Nakajima*, *Atsuro Ichigaya*, *Masaaki Kurozumi*
Kazuhisa Iguchi, *Yukihiro Nishida*, *Eisuke Nakasu*, *Seiichi Gohshi*

NHK Science and Technical Research Laboratories
1-10-11, Kinuta, Setagaya-ku, Tokyo 157-8510, JAPAN
sakaida.s-gq@nhk.or.jp

ABSTRACT

The Super Hi-Vision (SHV) developed by NHK is an ultrahigh-definition video system with 4000 scanning lines. Its video format comprises 7,680×4,320 pixels, which is 16 times the total number of pixels of HDTV (High Definition Television) and the frame rate is 60 Hz with progressive scanning. It has been designed to give viewers a strong sense of reality. In order to make the system suitable for practical use such as broadcasting services, a high-efficiency compression coding system is necessary. Therefore, we have developed several Super Hi-Vision codec systems based on MPEG-2 and AVC/H.264 video coding standards. SHV video signals are converted to sixteen HDTV signals and each video signal is processed by HDTV codec. The codec systems share some characteristics such as the frame synchronized encoding function. In this paper, details of these codec systems are described and transmission experiments using the codec are introduced.

***Index Terms*—** Super Hi-Vision, Video coding, MPEG, H.264, AVC

1. INTRODUCTION

Super Hi-Vision consists of an extremely high-resolution imagery system and a super surround multi-channel sound system [1]. Its video format comprises 7,680×4,320 pixels (16 times the total number of pixels of HDTV) and a 60-Hz frame rate with progressive scanning. It employs a 22.2 multi-channel sound system (22 audio channels with two low frequency effect channels). It has been designed to give viewers a strong sense of reality. The final goal of our research and development on SHV is to deliver super realistic image and sound to viewers' homes. When SHV becomes a broadcasting system, we will be able to use it for many purposes, such as archives and medical use. It is an important responsibility to preserve all of humankind's invaluable art, legacies and culture by recording using this ultimate system for future generations.

NHK has developed SHV cameras, projectors, disk recorders and audio equipment. Several SHV programs have been produced using these devices, and they have attracted many visitors at events such as the 2005 World Exposition in Aichi, Japan, as well as NAB2006 in Las Vegas, USA [2] and IBC2006 in Amsterdam, Netherlands [1].

In order to achieve our final goal of broadcasting SHV programs to homes, a new high-efficiency compression coding system is necessary. Therefore, we have developed several SHV codec systems that are based on MPEG-2 and AVC/H.264 video coding standards. It is necessary to conduct transmission tests and to solicit public opinion to encourage many people to recognize SHV as a broadcasting service. However, this cannot be done without an SHV codec. A real-time SHV codec will make people think of it as a real broadcasting system. For this reason, we need to develop the codec as soon as possible. In this paper, the outline of our SHV system and details of the SHV codec systems are described, and transmission experiments using the codec systems are introduced.

2. SUPER HI-VISION SYSTEMS

Table 1 lists the specifications of the SHV system, comparing them with those of HDTV. SHV is now the highest resolution TV system available. Basic parameters of the SHV system are designed to enhance the viewers' visual experience. NHK has developed key equipment for SHV broadcasting systems, such as cameras, display systems, audio systems and disk recorders. SHV requires an imaging device and display device with 32 million pixels; however, integrated devices with such huge resolution are not yet available. Thus, the SHV camera uses four panels with 8 million pixels each for green 1 (G1), green 2 (G2), red (R) and blue (B) channels, using the pixel-offset method to increase the effective number of pixels both horizontally and vertically. Figure 1 shows the pixel-spatial-sampling arrangement in this method.

Table 1 Specifications of SHV and HDTV

Specifications	SHV	HDTV
Number of pixels	7680 x 4320	1920 x 1080
Aspect ratio	16:9	16:9

Standard viewing distance (H: height of screen)	0.75H	3H
Standard viewing angle (horizontal)	100 deg.	30 deg.
Audio	22.2ch	5.1ch

Figure 2 shows the SHV roadmap for the future. NHK Science and Technical Research Laboratories (STRL) started the SHV study in 1995, and made steady progress towards its practical use within the first decade. We expect to launch SHV broadcasting by 2025. The plan is to deliver SHV to homes through a cost-effective network where it will then be recorded onto a home-use receiver. We anticipate that experimental SHV broadcasts will start in 2015 using a 21-GHz-band satellite, which is a potential delivery media for high-bit-rate transmissions. The display for SHV is another important subject. The widespread usage of large-size, high-resolution flat panel displays is remarkable. We assume that SHV displays for home use will be either a 100–200-inch large screen display or an A3-sized handheld-type paper-thin display with extremely high resolution. The SHV system has the potential for use in various applications in addition to broadcasting, for example, art, medical use, security and monitoring. In-theater presentation of sports events, concerts, etc. will be implemented before the broadcasting stage. The SHV systems can also be used in non-theater environments such as for advertisements, image archive materials and background images for program production.

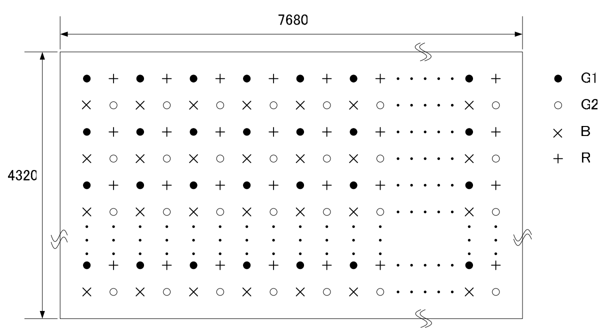


Figure 1 Pixel-spatial-sampling arrangement in the pixel-offset method

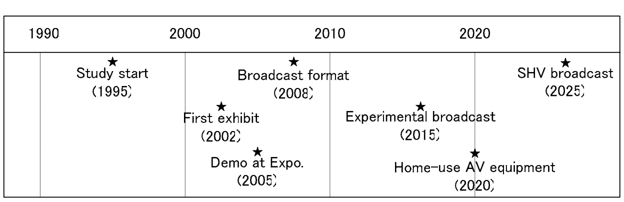


Figure 2 Roadmap toward the realization of SHV broadcasting

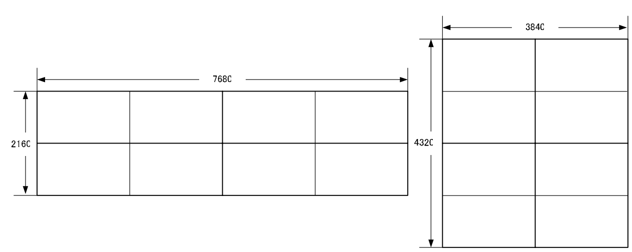
3. CODEC SYSTEMS

For realization of SHV broadcasting services to homes, video codec systems have been developed for efficient transmission and recording of SHV signals. The codec system consists of a video format converter, a video codec, and an audio codec. In this section, the video format converter and video codec systems are discussed.

3.1. Video format converter

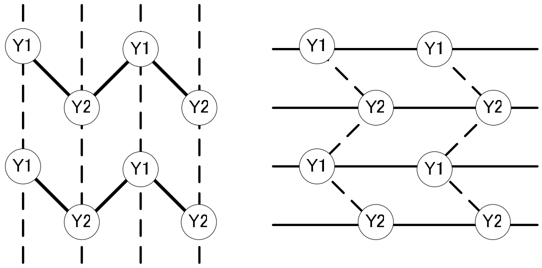
In order to encode the SHV signals in YUV format, the original SHV signals are converted for ease of handling. The video format converter converts the 7680×4320 (G1, G2, B and R) format from/into the sixteen 1920×1080/30 Psf (Progressive segmented frame) (Y/C 4:2:2, HD-SDI) images, where the SHV image is divided spatiotemporally. The color format conversion is formulated by (1), which is based on ITU-R Rec. 709. As previously mentioned, the current SHV signals are formatted by the pixel-offset method. Each signal component (G1, G2, B, and R) is a quarter of the size of SHV and is arranged in a pixel shift position as shown in Fig. 1. Since there are two luminance signals Y1 and Y2 after the video format conversion, the number of meaningful pixels becomes equal to half the SHV gross area 7680×4320. Therefore, in the case of conversion to the sixteen HD-SDI signals in practice, SHV signals should be spatially divided into eight parts and temporally into two parts. Spatial division should have two modes: (a) four horizontal parts and two vertical parts ("H" division); and (b) two horizontal parts and four vertical parts ("V" division), as shown in Fig. 3. The relationship of neighboring pixels of the YCbCr signal is drawn as solid (horizontal direction) and broken (vertical direction) lines in Fig. 4. Temporal division should be either two 60i signals or two 30Psf signals, as depicted in Fig. 5.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.7152 & 0 & 0.0722 & 0.2126 \\ 0 & 0.7152 & 0.0722 & 0.2126 \\ -0.1927 & -0.1927 & 0.5000 & -0.1146 \\ -0.2271 & -0.2271 & -0.0458 & 0.5000 \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \\ B \\ R \end{bmatrix} \quad (1)$$



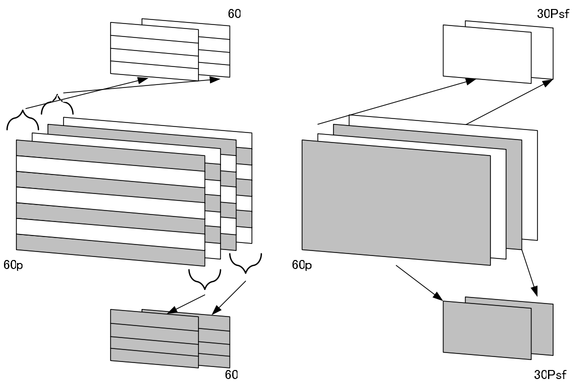
(a) "H" division (b)"V" division

Figure 3 Spatial division of SHV



(a) “H” division (b)”V” division

Figure 4 Relationship of pixels



(a) 60i (b) 30Psf

Figure 5 Temporal division of SHV

In order to decide the division mode, some preliminary experiments were conducted. Encoding experiments were performed by MPEG-2 HDTV codec systems on four combinations of spatial and temporal division of SHV signals using SHV images. The results showed that the combination of “H” and 30Psf division outperforms other combinations; hence, “H”-30Psf division was used for the design of the codec system. Figure 6 shows a diagram of the total codec system.

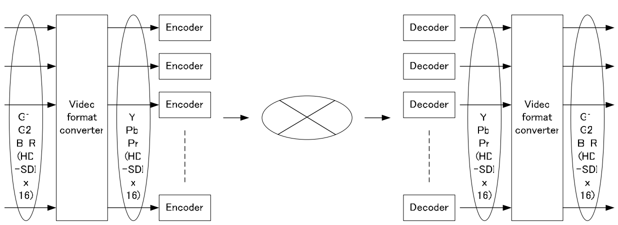


Figure 6 Diagram of codec system

3.2. MPEG-2 based codec

Since it was necessary to realize the SHV codec system as early as possible, we selected the MPEG-2 codec with its proven technology as the base system for the first version of the SHV codec [1][3]. The MPEG-2 based video codec consists of four sub-codecs for 3840×2160 images. A sub-codec contains four single-unit MPEG-2 HDTV codecs and a multi-channel frame synchronizer. To compress large images that exceed the resolution of HDTV by using MPEG-2 video coding, it is necessary to divide the SHV image into multiple HDTV units. The HDTV coding conforms to MPEG-2 Main Profile or 4:2:2 Profile @ High Level. Since motion vectors in higher resolution images are often significantly larger than those in lower resolution images, it is necessary to implement motion estimation with a wider search range in the encoder. The search vector range of the developed encoder achieves +/− 211.5 pixels horizontally and +/− 113.5 lines vertically, exceeding existing common encoders. The total coding delay is about 650 ms, including the encoder, the decoder and the video format converters. The main specifications of the MPEG-2 codec system are listed in Table 2 and its appearance is depicted in Fig. 7.

Table 2 Specification of MPEG-2 based codec system

Video	1920 x 1080/30 Psf x 16
	MPEG-2 4:2:2/Main profile
	180 – 600 Mbps
Audio	24 ch (AES3 x 12)
	PCM (48 kHz, 24bit), 28Mbps or Dolby-E, 7Mbps
TS I/F	MPEG-2 TS (DVB-ASI) x 4
	Max. 640 Mbps

Figure 7 View of MPEG-2 based codec system

The multi-channel sound signals can be transmitted in the form of uncompressed linear PCM (pulse code modulation) (48-kHz sampling, 24 bits/sample). In addition to the PCM, Dolby-E codecs with a compression ratio of 1:4 are also equipped. They handle 24 audio signal channels for the 22.2 multi-channel sound system. The coded video and audio signals are multiplexed into four MPEG-2 transport stream (TS) signals interfaced via DVB-ASI.

A TS recording device has also been developed, which supports the storage of long program material. The storage capacity is 1.2 Tbytes, which enables storage of a 4.5-hour-long program when coded at a bit-rate of 600 Mbps.

In the case of using public IP networks for transmission of the encoded SHV signals, consideration must be given to jitter and time delay depending on the transmission path. To synchronize the four TSs generated by the sub-encoders, the system manages the timing of each video frame by means of a time code and temporal reference in the GOP header of the MPEG-2 video stream. The sub-encoders communicate with each other via Ethernet. A master sub-encoder controls the start timing of all the sub-encoders. At the decoder, the master sub-decoder adjusts the display timing of all the sub-decoders and accounts for transmission delay by referring to the time code and temporal reference. The decoder can cope with the relative delay in the four TSs within 15 video frames. All HDTV decoders in sub-decoders work synchronously using black burst as a reference signal.

3.3. AVC/H.264 based codec

For the purpose of achieving very low bit-rate coding for SHV with high-quality images, a new codec system based on the AVC/H.264 coding scheme has been developed. AVC/H.264 is currently the most efficient standard video coding scheme, and is widely used in various applications such as broadcasting small images for mobile reception or HDTV services via a satellite network [4]. Since there are no AVC/H.264 codecs for images as large as SHV, sixteen HDTV AVC/H.264 codecs are used to construct SHV codecs similar to the MPEG-2 based codecs.

Each HDTV AVC/H.264 codec conforms to Main Profile @ L4 and will handle High Profile in the future. The encoder consists of three FPGAs and one DSP; therefore, the encoding process can be modified by replacing the software. One HDTV frame is divided into four slices and each slice is processed in parallel. Motion estimation, which is conducted on FPGA chips, has two phases: the first is pre-motion estimation, which is a rough prediction on one whole HDTV frame with two-pixel precision; the second is precise estimation on each of the four slices with quarter-pixel precision. The DSP chip, used mainly for rate control, administers the entire HDTV encoding processing module. The HDTV encoder is 1 rack unit (RU) in size and has DVB-ASI output. Each HDTV decoder is 1 RU in size and has the specifications for professional use.

Frame synchronization of the sixteen output images is the most important issue of the system, therefore, a new synchronization mechanism was developed in which one of the sixteen encoders becomes a master and the other fifteen encoders become slaves. In order to synchronize PTS/DTS and PCR for MPEG-2 TS of the output streams, all encoders share the same system date. The master encoder sends to all slaves a “start” hardware signal and 27-MHz clock, so that all encoders’ date counters increment at the same rate. GOP synchronization is also realized. All encoders generate an I frame when more than N encoders detect a scene cut change. The value N is programmable, and is usually set to 9. When an encoder generates an I frame independently from the others, it will generate the next I frame at the beginning of GOP to stay synchronized with the others.

The statistical properties of the SHV source signals are not effectively utilized in the current HDTV encoders. A future task is to share the SHV source information between the encoders to improve compression performance further.

The HDTV AVC/H.264 codec has been developed jointly with Fujitsu Laboratories Ltd.

4. DEMONSTRATIONS AND EXPERIMENTS

On March 14, 2006, NHK carried out an experimental transmission of SHV using the MPEG-2 coding system at bit-rates of 300 and 640 Mbps via gigabit IP networks over a distance of 17.6 km. The IP transmission system was also successfully demonstrated at NAB 2006 in the USA from April 24 to 27 [1], and IBC 2006 in the Netherlands from September 8 to 12, 2006 [2]. Every New Year’s Eve, a popular music show is broadcast from the NHK Hall. On December 31, 2006, NHK conducted an experiment on live relay broadcast from Tokyo to Osaka for public viewing.

The new AVC/H.264 codec systems at 128 Mbps were demonstrated at NHK STRL open-house in May 2007 and NHK is planning to conduct a transmission experiment in August 2007.

5. CONCLUSION

Significant progress has been made toward implementing the practical use of SHV, particularly the development of SHV codecs and transmission systems for contribution and distribution purposes. These systems can be utilized for various events such as cooperative museum exhibitions, live relay of sports events on a global scale, and public viewing of SHV programs.

NHK will continue its tireless research and development efforts along with the roadmap aiming at launching SHV broadcasting in 2025.

6. REFERENCES

[1] E. Nakasu, Y. Nishida, M. Maeda, M. Kanazawa, S. Yano, M. Sugawara, K. Mitani, K. Hamasaki and Y. Nojiri, “Technical development towards implementation of extremely high resolution imagery system with more than 4000 scanning lines,” *IBC2006, Amsterdam, Sep. 11, 2006.*

[2] M. Maeda, Y. Shishikui, F. Suginoshita, Y. Takiguchi, T. Nakatogawa, M. Kanazawa, K. Mitani, K. Hamasaki, M. Iwaki and Y. Nojiri, “Steps Toward the Practical Use of Super Hi-Vision,” *NAB2006 Proceedings, Las Vegas, Apr. 26, 2006.*

[3] Y. Nishida, A. Ichigaya, M. Kurozumi and E. Nakasu, “MPEG-2 codec for Super Hi-Vision,” *2006 IEICE General Conference, Tokyo, Mar. 24, 2006 (in Japanese).*

[4] G. J. Sullivan, “The H.264/MPEG-4 AVC video coding standard and its deployments status,” *VCIP2005, Beijing, Jul. 14, 2005.*