

フル HDTV リアルタイム 3D 映像エンコーダ装置の開発

岩崎 裕江 大西 隆之 佐野 卓
横張 和也 中島 靖之 小林 大祐
嵯峨田 淳 高橋 裕子 新田 高庸
上倉 一人

近年、高臨場感映像を実現する3Dカメラ、3Dディスプレイ、3D配信などの3D技術の研究開発が盛んである。我々は、高臨場感映像のリアルタイム配信を可能とするフルHDTVリアルタイム3D映像エンコーダ装置を開発した。本稿では、フルHDTVリアルタイム3D映像エンコーダ装置の構成およびフィージビリティについて示す。

Full HDTV Real-time 3D Video Encoder

Hiroe Iwasaki Takayuki Onishi Takashi Sano
Kazuya Yokohari Yasuyuki Nakajima Daisuke Kobayashi
Atsushi Sagata Yasuko Takahashi Koyo Nitta
Kazuto Kamikura

Recently, developments of 3D technologies such as 3D cameras, 3D displays, and 3D delivery are very important because of realistic video environment. We proposed the full HDTV real-time 3D video encoder system for real-time delivery. This paper presents full HDTV real-time 3D video encoder system configuration and its feasibility.

1. はじめに

近年、米国を中心に3D映画が急速に立ち上がり、臨場感のある3D映像への期待が高まっている。臨場感のある3D映像は、映画だけではなく、IPTVの3D放送やサッカーなどのスポーツイベントなどのリアルタイム配信（パブリックビューイング）、デジタルサイネージなど幅広い範囲に適用可能であり、利用価値の高い技術である。また、ITU-TやISO/IECの符号化グループにより、H.264 Stereo High Profile[1,2]が規格化され、3D関連技術の研究開発が盛んにおこなわれている。H.264 Stereo High Profileでは、既存のH.264との互換性を保持するために、マクロブロック以下のシンタックスは同一であるという大きな特徴を持っている。さらに、H.264 AVC 2チャンネルの符号量に比較し、左目画像（ベースビュー）と右目画像（非ベースビュー）の相関をとった場合、非ベースビューは平均20%~30%、最大43%ものビット量を削減できる[3]。

これに対し、我々は、高臨場感リアルタイム配信を目指し、H.264 Stereo High Profile規格に準拠したフルHDTVリアルタイム3D映像エンコーダ装置を我々が既に開発したH.264 High Profileエンコーダ[4]を用いて実現した。以下に、フルHDTVリアルタイム3D映像エンコーダ装置を実現するための課題を示す。

- H.264 Stereo High Profile によるビュー間予測：
H.264 Stereo High Profile のビュー間予測を用いた場合に、平均20%~30%の圧縮を実現することができる。リアルタイム配信では、ネットワークを介して伝送するため、圧縮効率の高いビュー間予測の実現は不可欠である。
- ベースビューおよび非ベースビューのストリームの同期：
同じ時刻に入力されたベースビューと非ベースビューを独立に符号化した場合、個々の時刻情報を用いると、表示タイミングのずれが発生する。このため、同時に入力されるベースビューと非ベースビューの時刻の同期を実現する必要がある。

これらの問題を解決した高臨場リアルタイム配信に適用可能なフルHDTVリアルタイム3D映像エンコーダの構成を提案し、そのフィージビリティを示す。

2. ビュー間予測実現方式

2.1 ビュー間予測と圧縮効率

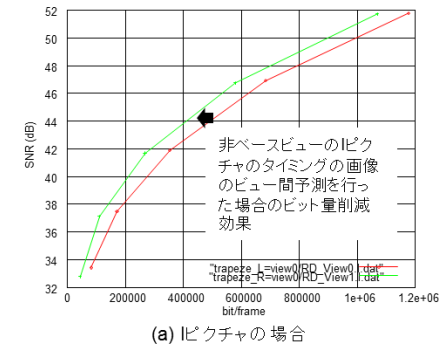
H.264 Stereo High Profile でのビュー間予測の効果を定量的に解析するために、H.264/AVCのリファレンスソフトウェアであるJM[5]を用いてシミュレーションした。ビュー間予測を用いた場合のピクチャタイプごとのビット量削減効果を図1(a)~(c)に示す。また、その時の符号化条件を表1、シミュレーション映像を図2に示す。本シミュレーションでは、評価画像として準備されている映像SDTV(720x480)のステレオ映像を用いた。また、GOP構造は、放送の規定に従いN=15とした。M値については、現状、放送用機器で多く用いられている値を用いた。図2のシミュレーション映像を用いて、シミュレーションした後、ベースビューのピクチャタイプ別にビット量を合計しビット量を図1に示している。図1に示すように、ベースビューのIピクチャと同時に入力される非ベースビューのビュー間予測のビット量削減効果は極めて大きい。また、ベースビューのPピクチャやBピクチャと同時に入力される非ベースビューのビュー間予測のビット量削減効果は非常に小さい。このため、ベースビューのIピクチャと同時に入力される非ベースビューのピクチャのみビュー間予測を行う方法は極めて有効である。

2.1 ビュー間予測実現方式

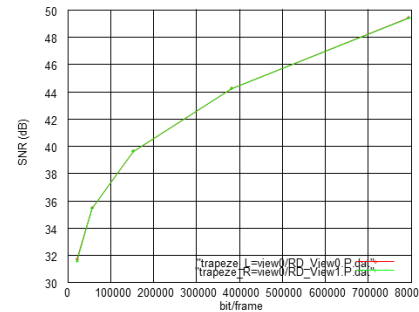
ベースビューがIピクチャと同時に入力される非ベースビューのみビュー間予測を行い、ベースビュー・非ベースビューが独立してH.264の符号化を行う場合の参照関係を図3に示す。図3に示すように、ベースビューと非ベースビューでは、15枚に1枚のベースビューのタイミングを除いて、独立に符号化を行うことができる。ベースビュー、非ベースビューを独立したH.264High Profileエンコーダを用いて、図3に示す参照関係によるビュー間予測を行う方法を図4に示す。視差探索は、画像の

表1 シミュレーション条件

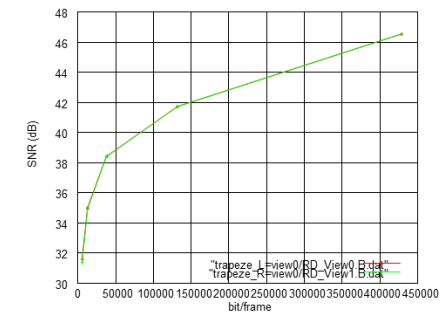
符号化方式	H.264 Stereo High Profile
ピクチャ構造	フィールド構造
GOP構造	ベースビュー：M=3/N=15 非ベースビュー：M=3
画像サイズ	720 x 480
フレーム数	300



(a) ピクチャの場合



(b) Pピクチャの場合



(c) Bピクチャの場合

図1 ピクチャタイプ別ビュー間予測のビット量削減効果



図2 シミュレーション映像(Trapeze)

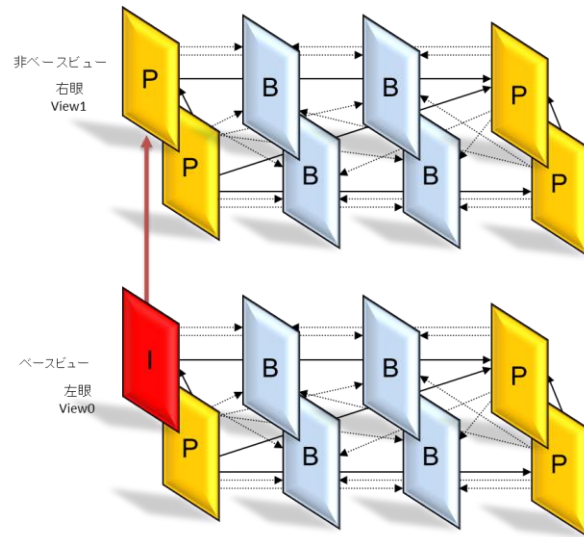


図3 ベースビューと非ベースビューの参照関係

動き探索と同様の処理で実現することが可能だが、その特性は、時間方向の動きの場合とは大きく異なる。このため、視差探索は、H.264High Profile エンコーダに映像を入力する前段で行う。視差探索した結果のベクトルは、ベクトル情報として、画像に重畳し、非ベースビューを符号化する H.264High Profile エンコーダに入力する。非ベースビューを符号化する H.264High Profile エンコーダは、画像に重畳されたビュー間予測ベクトルを用いて符号化する。非ベースビューで参照画像とするベースビューの画像 (図3の赤色フレーム) のローカルデコード画像を一致させるために、ベースビューと非ベースビューを独立に符号化する H.264High Profile エンコーダで符号化制御情報や時刻情報を共有する。また、非ベースビューのエンコーダは参照画像に使用したビットストリームを除いて、多重化部へ出力する。多重化部では、独立した H.264High Profile エンコーダから出力されるトランスポートストリームを再多重化して出力する。

3. 3D 映像エンコーダの実装

3.1 3D 映像エンコーダの構成

フル HDTV リアルタイム 3D 映像エンコーダ装置の構成を図5に示す。図5に示すように、本 3D 映像エンコーダは、1920x1080i のフル HDTV を H.264High Profile で符号化可能な H.264 符号化モジュール、Host CPU、FPGA、メモリで構成される。各 1920x1080i の H.264 符号化モジュールは、H.264 Stereo High Profile の規格でマクロブロック以下のシンタックスが同じであることを利用し、既開発の H.264 High Profile エンコーダ[4]モジュール(SARA モジュール)を流用した。ベースビューと非ベースビューで共有するベースビューの I ピクチャ画像は、I ピクチャの符号化に必要な数個程度の符号化パラメータのみ Host CPU を介してデータ共有し、それぞれの H.264 符号化モジュールで独立に符号化する。また、Host CPU を介してベースビューの時刻情報を非ベースビューへ出力することにより、ベースビューと非ベースビューのストリームの同期を実現する。

FPGA は、視差探索部、視差ベクトル重畳部、多重化部から構成され、複数のフレームメモリに接続され、FPGA の本フレームメモリを介して、非ベースビューで参照されるベースビューの原画像を非ベースビューの入力画像として出力し、H.264High Profile エンコーダモジュールの入力画像データの制御を行っている。また、FPGA では、各 H.264High Profile エンコーダモジュールから出力されるトランスポートストリームを再多重化する機能を具備している。

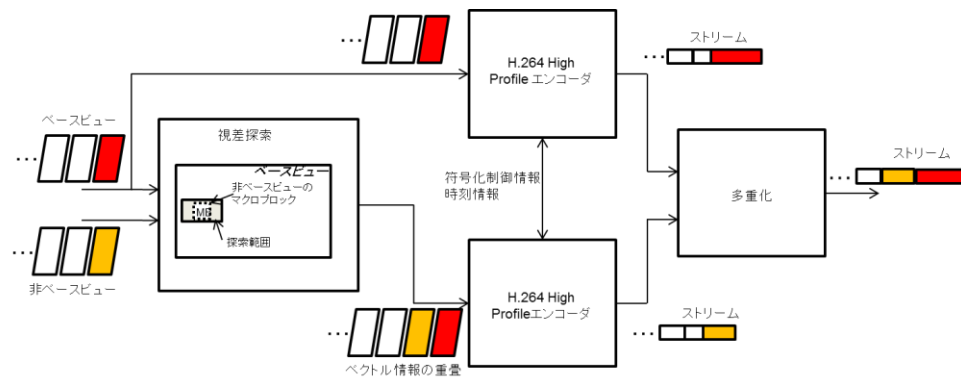


図4 ビュー間予測実現方式

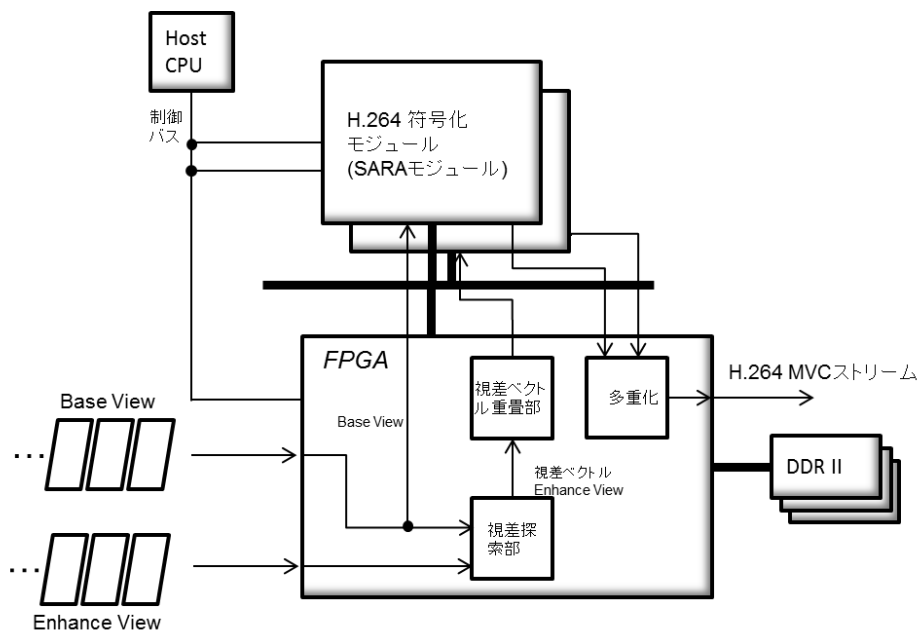


図5 3D映像エンコーダ装置の構成

3.2 3D映像エンコーダの諸元

本3D映像エンコーダ装置の外観写真と諸元を図6及び表2に示す。本3D映像エンコーダ装置は、HD-SDI入力を2チャンネル具備し、放送プロ向けに開発されたH.264符号化モジュールを使って、H.264 Stereo High Profile ストリーム(H.264 Stereo High Profile)をDVB-ASIに出力する。入力可能な画像サイズは、1080iおよび720Pである。GOP構造は、N=15/M=3を使用し、1Uで実現している。



図6 3D映像エンコーダ装置の写真

4. おわりに

本稿では、高臨場リアルタイム通信を目指したフルHDTVリアルタイム3D映像エンコーダの構成およびそのフィージビリティについて示した。今後は、さらなる視差探索の精度の向上および符号化制御などのチューニングなどにより、圧縮効率の向上を目指す。

表2 3D映像エンコーダ装置諸元

入力	HD-SDI x 2
出力	DVB-ASI
入力画像	1080i x 2チャンネル または 720P x 2チャンネル
符号化方式	H.264 Stereo High Profile
GOP構造	N=15 / M=3
装置サイズ	1U

参考文献

- 1) ITU-T and ISO/IEC JTC 1, Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496-10(MPEG-4 AVC), 2010.
- 2) A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, Joint draft 8 of multiview video coding, Hannover, Germany, Joint Video Team(JVT) Doc. JVT-AB204, Jul. 2008.
- 3) A. Vetro, T. Wiegand, G. J. Sullivan: Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard, Proceeding of the IEEE, Vol. 99, No. 4, pp. 1-17(2011)
- 4) K. Nitta, M. Ikeda, H. Iwasaki, T. Onishi, T. Sano, A. Sagata, Y. Nagajima, M. Inamori, T. Yoshitome, H. Matsuda, R. Tanida, A. Shimizu, K. Nakamura and J. Naganuma An H.264/AVC High 422 Profile and MPEG-2 422 Profile Encoder LSI for HDTV Broadcasting Infrastructures, 2008 IEEE Symposium on VLSI Circuits, pp.106-107, 2008.
- 5) Y. Chen, P. Pandit, S. Yea, and C. S. Lim, Draft reference software for STEREO HIGH PROFILE(JSTEREO HIGH PROFILE 6.0) London, U.K., Joint Video Team (JVT) Doc. JVT-AE207, Jul. 2009.