

MPEG2 MVP@HL による HDTV 立体符号化における 左・右画像へのレート配分

今泉浩幸 酒井美和 蓑毛研 岩館祐一

NHK 放送技術研究所

〒157 東京都世田谷区砧 1-10-11

2眼式立体テレビ符号化用に標準化された MPEG2 マルチビュープロファイル (MVP) 方式を立体 HDTV シーケンスに適用した。MVP では左眼用画像をメインプロファイル (MP) で符号化し、右眼用画像を動き補償予測と視差補償予測を用いて符号化する。従来の MP 方式を左右のシーケンスに独立に適用するサイマルキャスト方式と MVP 方式を符号化 SNR により比較した。さらに、MVP 方式においては、左右シーケンスへのビットレート配分を適応的に制御することにより左右の画像の平均 SNR 値を改善できることを示した。

Bitrate allocation for left and right view sequences on stereoscopic HDTV coding using MPEG2 MVP@HL

Hiroyuki IMAIZUMI, Miwa SAKAI, Ken MINOMO, Yuichi IWADATE
NHK Science and Technical Research Laboratories

This document reports results of computer simulation on the performance of a stereoscopic HDTV coding using the MPEG2 Multi-View Profile (MVP). The performance of MVP is compared with a conventional Main Profile (MP) simulcast approach in terms of SNR of decoded images. Bitrate allocation for left and right view sequences is examined. It is observed that an adaptive bit allocation depending on image contents improved the average SNRs of decoded left and right images.

1. はじめに

2眼式立体テレビ画像の符号化方式 MPEG2 マルチビュープロファイル (MVP: Multi-View Profile) [1]の MPEG2 ビデオパート[2]への追加が 1996 年 MPEG 9月会合で承認された。MVPはテンポラルスケラビリティ機能を利用して、2眼式立体画像の左眼用シーケンスを基本レイヤで、右眼用シーケンスを拡張レイヤで符号化する。スケラブル符号化を利用することで、基本レイヤのビットストリームのみを復号すれば、左眼用のシーケンスが復号できることになる。したがって、MVPのビットストリームはMP(Main Profile)復号器での復号の両立性が確保されていることになる。さらに、拡張レイヤにおける右眼用シーケンスの符号化では、従来の動き補償予測に加えて、視差補償予測(基本レイヤ画像からの予測)を行うことにより符号化効率の改善を図っている。

これまで、MVPの標準化作業の中で、ML(Main Level: 標準テレビの画素数に対応)のステレオシーケンスについて、左・右の画像を独立にMPで符号化するサイマルキャスト方式に対する MVP方式の優位性が、符号化誤差電力の比較、画質主観評価により検証されている[3]。しかしながら、MVP方式のビットレートと画質の関係や立体HDTVへの適用については検討の余地が残されている。

本稿では、はじめに MVPによる立体テレビ符号化方式について紹介し、次に MVPを2眼式立体HDTVシーケンスの符号化に適用した場合について、サイマルキャスト方式との比較を行う。さ

らに、左眼・右眼用シーケンスへのビットレート配分について検討し、最適なレート配分は画像内容に依存し、画像内容により適応的にレート配分を行うことで画質の向上が期待できることを示す。

2. MPEG2 MVP 符号化方式

MVP に対応した符号化器の構成例を図 1 に示す。基本レイヤで伝送する左眼用のビットストリームは通常の MP 復号器によっても復号可能とするため MP 符号器によって符号化される。一方、拡張レイヤで伝送する右眼用のシーケンスは符号化効率改善のため、動き補償予測と視差補償予測を適応的に用いて符号化される。

MVP における予測方式の例を図 2 に示す。基本レイヤの予測方式は MP の予測方式と同一である。拡張レイヤではフレーム内符号化を行う I ピクチャ、1 方向予測を利用できる P ピクチャ、2 方向予測を利用できる B ピクチャを使用することができる。P ピクチャは基本レイヤの同一時刻に表示されるピクチャ(視差補償予測)または拡張レイヤで直前に符号化されたピクチャ(動き補償予測)のどちらかをピクチャ単位に選択して参照することができる。B ピクチャは基本レイヤの同一時刻に表示されるピクチャ(視差補償予測)及び拡張レイヤで直前に符号化されたピクチャ(動き補償予測)の両方または一方をマクロブロック単位に選択して参照することができる。ただし、基本レイヤで符号化されるピクチャが GOP(Group of Pictures: 1 個以上のピクチャから構成される符号化の単位で、GOP の最初のピクチャは I ピクチャを使用する。したがって、GOP 単位に復号を開始することができる。)の最初のピクチャである場合

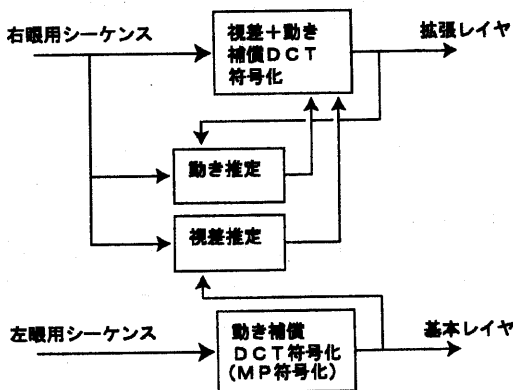


図1 MPEG2 MVP 符号化器の構成例

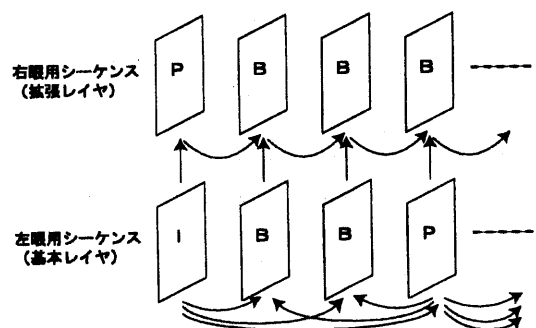


図2 MVPの予測方式 (基本レイヤ:M=3)

には、拡張レイヤの対応するピクチャはIピクチャまたは基本レイヤを参照するPピクチャとし、GOP内で閉じた予測参照構造とすることが規定されている。

図2の例では拡張レイヤのGOPの最初のピクチャは基本レイヤを参照するPピクチャである。この例のように拡張レイヤではIピクチャつまりフレーム内符号化を用いない構成が可能であり、Iピクチャに多くのビットを割り当てる必要がない。また、ピクチャタイプにより発生ビット量が大きく変動することがなく、レート制御が容易である。

MP復号器との両立性を持ったストリームを生成するスケーラブル符号化方式はMPEG2のテンポラルスケーラビリティ機能により実現されている。テンポラルスケーラビリティは時間解像度の階層符号化を実現するための機能であり、たとえば、基本レイヤまで復号するとインターレース画像が得られ、さらに拡張レイヤまで復号するとプログレッシブ(ノンインターレース)画像が得られるといった符号化を想定して用意されている。しかしながら、このような時間解像度の階層符号化を行うプロファイルは、現在のところ規定されておらず、唯一、MVPだけがテンポラルスケーラビリティ機能を利用するプロファイルとなっている。

3. 計算機シミュレーション

3.1 シミュレーション仕様

シミュレーションに用いた符号化アルゴリズムは、ISO/IEC 13818-5 MPEG2ソフトウェアシミュ

レーションパートの仕様を基本とし、拡張レイヤの符号化では予測フレームに基本レイヤの画像も参照できるように変更した。ピクチャ構造はフレーム構造とし、GOP内のフレーム数は15とした。拡張レイヤの予測構造は図2の例にしたがってGOPの最初のフレームは基本レイヤを参照するPピクチャ、GOPの残りのフレームはBピクチャとした。なお、基本レイヤでは3フレーム毎にIピクチャまたはPピクチャ(N=15, M=3)とする予測構造を用いた。拡張レイヤでのレート制御では、ピクチャタイプ(Pピクチャ、Bピクチャ)による割り当てビット量の重み付けは行わず、各ピクチャの符号化開始時にGOPに残されたビット量を残りのフレームで等分した値をそのピクチャの目標符号化ビットレートとすることとした。

MVPではHL(High Level: HDTVの画素数に対応)でのVBV(Video Buffering Verifier: 符号器のレート制御を規定するために仮定された復号器モデル)の最大バッファサイズが基本レイヤで9787248ビット、基本レイヤを含み拡張レイヤまでで15898480ビットと規定されている。このシミュレーションでは簡単のため基本レイヤと拡張レイヤのストリームを独立にレート制御することとし、拡張レイヤのストリームの拘束条件はVBVバッファサイズを6111232ビット(拡張レイヤまでの最大バッファサイズと基本レイヤの最大バッファサイズの差)として規定した。この場合には、基本レイヤと拡張レイヤでそれぞれにVBVバッファ条件を満足するように制御すれば、規定の拘束条件を満足できることになる。拡張レイヤでは低いビットレートでの符号化画像を得るために、よ

表1 予測ベクトルの探索範囲 (画素数)

基本レイヤ	拡張レイヤ
(動き補償予測) フレーム間距離1: 水平±31、垂直±31 フレーム間距離2: 水平±63、垂直±31 フレーム間距離3: 水平±127、垂直±31	(視差補償予測) 同時刻: 水平±63、垂直±7 (動き補償予測) フレーム間距離1: 水平±31、垂直±31
StreetOrgan (右眼用) 10Mbps および Ball-Star (右眼用) 10, 12, 14, 16Mbps については、 フレーム間距離1: 水平±31、垂直±31 フレーム間距離2: 水平±7、垂直±7 フレーム間距離3: 水平±63、垂直±31	StreetOrgan 8, 10Mbps については、 (視差補償予測) 同時刻: 水平±63、垂直±7 (動き補償予測) フレーム間距離1: 水平±7、垂直±7

り広い範囲の量子化スケールを利用できる $q_scale_type = 1$ により量子化スケールの制御を行った。また、比較のために求めた右眼用シーケンスのMP符号化においても $q_scale_type = 1$ とした。

動き補償予測および視差補償予測の探索範囲を表1に示す。探索精度は $1/2$ 画素精度である。いくつかのビットレート及びシーケンスでは動き・視差ベクトルの伝送に割り当てられるビット量が大きく、所望のビットレートまたは画質が得られない場合があり、狭い探索範囲を用いた。具体的には、ベクトルの伝送ビット量を低減するためB-ピクチャにおいて、基本レイヤではフレーム間距離2の参照フレームからの予測ベクトルの探索範囲、拡張レイヤでは動き補償予測ベクトルの探索範囲を狭くし、2つのベクトルを使用する内挿予測モードの割合を小さくする手法を用いた。

3. 2 評価画像

シミュレーションに用いた画像を図3に示す。2台のHDTVカメラにより撮像した立体HDTVシーケンスであり、各シーケンスの30フレームについて符号化した。(a) *Balloons* は背景が静止した建物の壁や樹木であり、その前景を大勢の子供たちが走り抜けていくシーンである。(b-2) *StarSailor* はCGで作成した銀河とカメラで実写した自然画像を合成したシーンで、面積的には小さいものの実写部分については光の反射により左右の画像で輝度差が大きいこと、平均的には輝度が低く、暗いシーンであることが特徴である。今回のシミュレーションでは30フレームの最初のフレームが *Balloons* の最初のフレーム、最後のフレームが *StarSailor* の最後のフレームとなるフェード画像 (b-1, b-2) *Ball-Star* を作成し評価に使用した。フェード画像は、時間的に平均輝度が変化するため動き補償予測の難しいシーケンスの例として用意した。(c) *StreetOrgan* は背景が建物の壁の近景、その前を右方向へ横切る人物と左方向へのカメラのパンニングが含まれており、動き領域の割合が大きいシーケンスである。フォーカスが少女の位置にあり、前景の人物はややボケている。

3. 3 シミュレーション結果と考察

(1) MVP方式とサイマルキャスト方式の比較

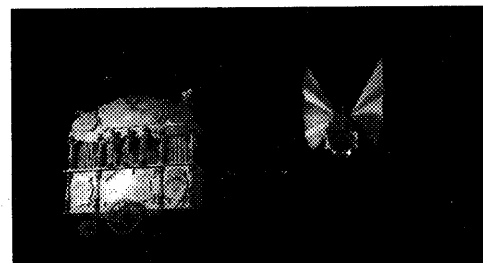
図4にMP@HLで符号化した左眼用シーケンスのビットレートと符号化SNRの関係を、図5にサイマルキャスト方式で使用するMP@HLで符号化



(a) Balloons [Frame 0] (Ball-Star [Frame 0])



(b-1) Ball-Star [Frame 14]



(b-2) Ball-Star [Frame 29] (StarSailor [Frame 29])



(c) StreetOrgan [Frame 0]

図3 立体HDTV評価画像

した右眼用シーケンスのビットレートと符号化 SNR の関係を示す。SNR 値は 30 フレームの符号化誤差電力の平均値である。左眼用の符号化画像は MVP@HL の基本レイヤにおける符号化画像と同一である。右眼用シーケンスの符号化では、2 乗予測誤差最小となる予測モードを選択する手法を用いているが、評価画像 Ball-Star と StreetOrgan については動きベクトルの伝送に必要なビット量が多くなり、すべてのフレーム間距離からの予測ベクトルの探索範囲を水平±7画素、垂直±7画素と狭くしても 8Mbps 以下のストリームを作成することができなかった。

図 6、図 7、図 8 に拡張レイヤで符号化した右眼用シーケンスのビットレートと符号化 SNR の関係を示す。基本レイヤで符号化した左眼用シーケンスのビットレートをパラメータとした。また、比較のために右眼用のシーケンスを MP@HL で符号化した場合、すなわち左眼・右眼用のシーケンスを独立にサイマルキャスト方式により伝送する場合の符号化 SNR を実線で示した。右眼用シーケンスの符号化に用いるビットレートを低くするにつれて MVP 方式による符号化 SNR の改善効果が顕著となる。特に、StreetOrgan で MVP 方式による改善が大きい。これは動き領域の割合が大きく、視差補償予測が有効に作用したためと考えられる。

一方、画像内容による違いはあるものの右眼用シーケンスの符号化ビットレート 16Mbps 以上ではサイマルキャスト方式の方が高い符号化 SNR 値を示している。I-ピクチャに十分なビットが割り当てられる場合には、視差・動き補償による両方向予測よりも過去と未来から予測する両方向動き補償予測の予測精度が高いためと予想される。

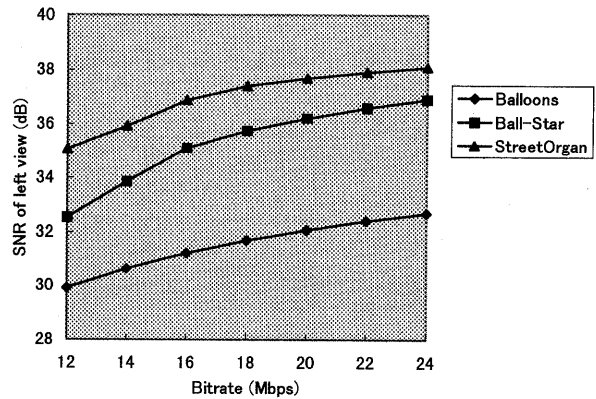


図 4 MP@HL (MVP 基本レイヤ) による左眼用画像の SNR

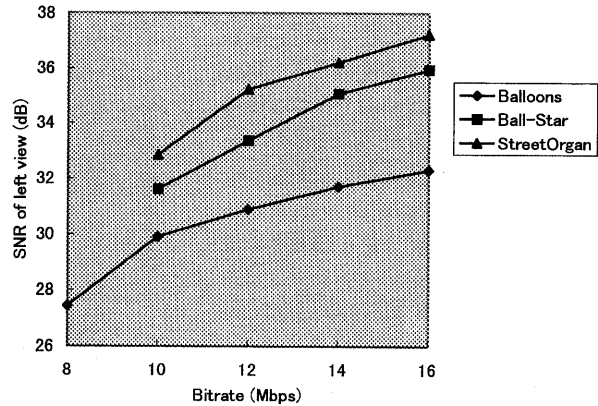


図 5 MP@HL (サイマルキャスト) による右眼用画像の SNR

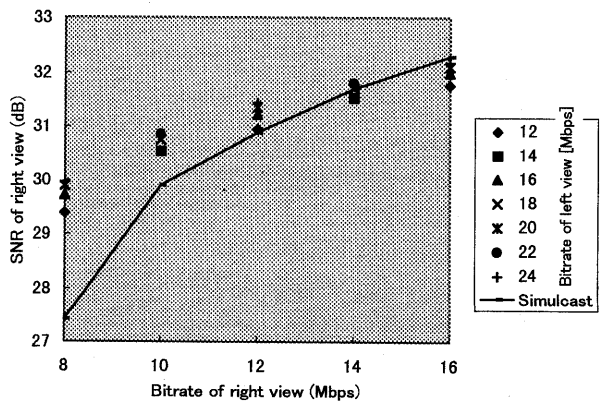


図 6 右画像のビットレートと SNR (Balloons)

(2) 左画像ビットレートが右画像 SNR 値へ与える影響について

符号化する拡張レイヤのビットレート一定の条件で、左眼用シーケンスを符号化する基本レイヤのビットレートと右画像の符号化 SNR の関係に着目する。たとえば、図 9 は拡張レイヤのビットレート 12Mbps の場合について基本レイヤのビットレートと右画像の SNR 値の関係を示したものである。左画像のビットレートの右画像の SNR 値への影響はフェード画像 *Ball-Star* で大きくなっている。*Balloons* では、画像内容に静止領域が多く含まれるため、視差補償予測（基本レイヤを参照画像とする予測）を用いる予測モードが選択される割合が小さいと考えられ、左画像のビットレートの影響が小さくなっている。*StreetOrgan* については動き領域が大きく、視差補償が選択されやすいことから依存性が高くなると予想していたが、顕著な傾向は認められなかった。前景の人物の部分など解像度の低い領域があり、低いビットレートにおいても、左画像の符号化 SNR が高くなったことが原因と思われる。

(3) 左右画像への最適レート配分

基本レイヤと拡張レイヤに対するレート配分について検討する。「符号化した立体シーケンスを立体視した場合の画質について、主観評価における評点は、左・右のシーケンスをそれぞれ立体視せずに評価した場合の評点の平均値となる。」との報告[4]があり、ここでは、左・右のシーケンスの平均の符号化 SNR が高いときに立体視時の画質も高くなると仮定する。

図 10、図 11、図 12 はそれぞれ基本レイヤと拡張レイヤのビットレートの和を 24Mbps、28Mbps、32Mbps とした時の基本・拡張レイヤへのビットレートの配分と左右の画像の符号化誤差電力の平均値に相当する SNR 値との関係を示したものである。また、左・右そ

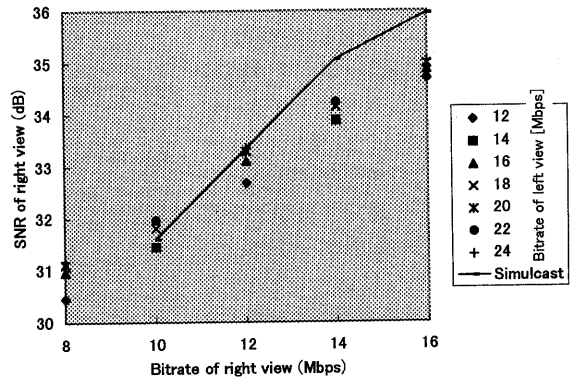


図 7 右画像のビットレートと SNR (*Ball-Star*)

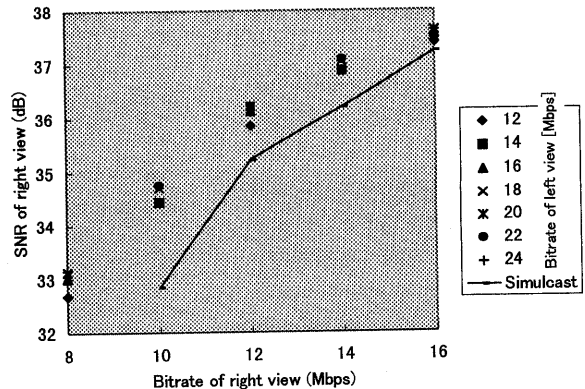


図 8 右画像のビットレートと SNR (*StreetOrgan*)

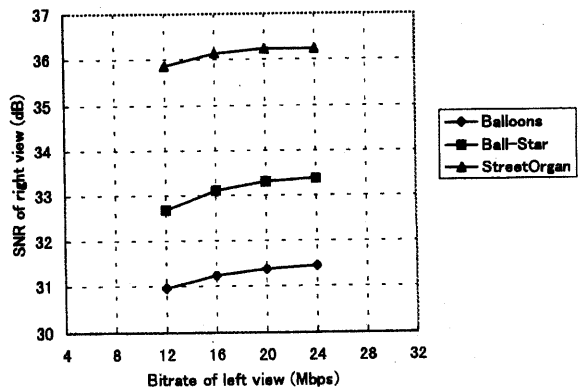


図 9 基本レイヤのビットレートと右画像の SNR

それぞれのシーケンスの符号化 SNR 値をあわせて示した。

トータルビットレート 28Mbps (図 11) においては、Balloons では (基本レイヤー-拡張レイヤ) のビットレートを (18Mbps-10Mbps) から (16Mbps-12Mbps) に設定した場合に SNR 値が最大となっている。左・右それぞれの SNR 値は (16Mbps - 12Mbps) 付近で等しくなっている。StreetOrgan では (16Mbps-12Mbps) の配分で左右の平均 SNR 値の最大が得られている。フェード画像 Ball-Star では検討したビットレートの組み合わせの中では、基本レイヤに割り当てるビットレートの割合を高くするほど左右平均の SNR 値が高くなるという結果が得られた。

フェード画像では時間的に平均輝度に変化するため動き補償予測の残差が大きくなり、他の画像と比較して、基本レイヤの画像を予測の参照画像とする視差補償予測が用いられる割合が高くなるため、右画像の SNR 値が左画像の SNR 値に依存しやすくなると考えられる。ただし、この最大値を与えるレート配分では左・右の SNR 値の差がかなり大きくなっており、主観的な画質評価による結果と一致するかどうかを調べる必要が残っている。

トータルビットレート 24Mbps (図 10)、32Mbps (図 12) の場合にも、左右の平均の SNR の最大値は、左・右それぞれの SNR 値が等しくなるレート配分に対して、やや基本レイヤ (左画像) へのレート配分を高くした配分で得られている。図 10 の StreetOrgan の場合のみこの傾向が明らかでなく、原因はよく分かっていないが、その他の場合には一般的な傾向と言える。

図 9 についての考察で示したように右画像の SNR 値は左画像の SNR 値に依存するため、左画像の SNR 値が右画像の SNR 値よりやや高くなるようにレート配分した場合に平均の SNR 値

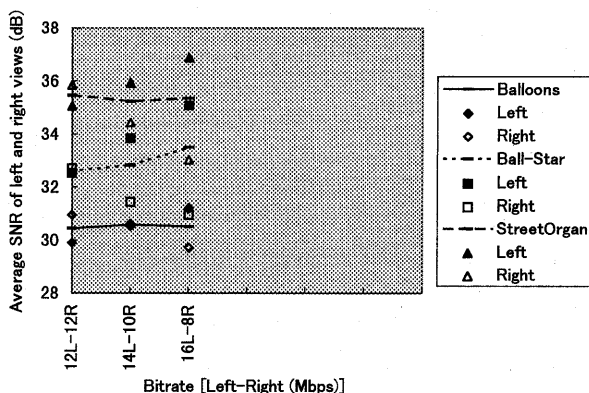


図 10 左右画像の平均 SNR (トータル 24Mbps)

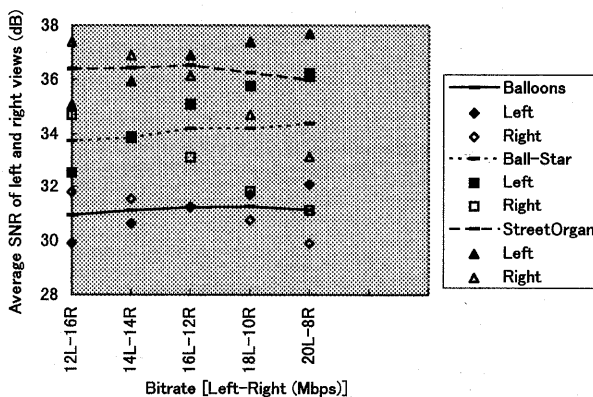


図 11 左右画像の平均 SNR (トータル 28Mbps)

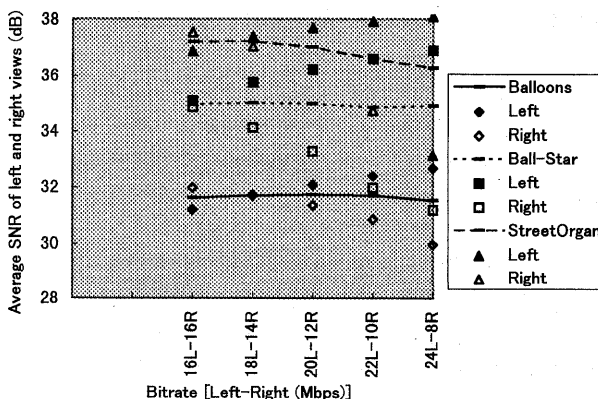


図 12 左右画像の平均 SNR (トータル 32Mbps)

が最大となると考えられる。この最大値を与えるレート配分は画像内容に依存する。したがって、画像内容に応じて基本・拡張レイヤへのレート配分を適応的に制御することにより、画質の改善が期待できる。

(4) 実験結果のまとめ

以上の立体 HDTV 評価画像を用いたシミュレーション実験から得られた結果をまとめると、

- ・ MVP 方式による拡張レイヤの符号化 SNR 値は、ビットレートが低くなるにしたがって、サイマルキャスト方式に対する優位性が顕著となる。
- ・ MVP 方式では、Iピクチャが必須でなく、レート制御が容易である。このため MP@HL ではレート制御が困難となる低ビットレートにおいても符号化が可能である。
- ・ MVP 方式において、左右の平均 SNR 値が最大となる基本・拡張レイヤへのレート配分は画像内容——たとえば、フェード画像であることや静止領域の割合など——に依存し、画像内容によってこのレート配分を制御することで左右の平均 SNR 値を改善できる。

となる。

4. おわりに

2 眼式立体 HDTV シーケンスの符号化において、左右の画像を独立に MP 方式で符号化するサイマルキャスト方式に対して MVP 方式が有利となる条件を示した。また、MVP 方式においては基本・拡張レイヤへのビットレート配分を制御することにより左右シーケンスの平均の SNR 値を改善できることを示した。得られた結果は、視差補償予測の有効な、逆に言えば、動き補償予測の有効でないシーケンスでは、より多くのビットを基本レイヤに配分する制御方式の有効性を示唆している。動き補償予測の有効でないシーケンスは、一般に符号化の難しいシーケンスと考えられ、このような場合に基本レイヤに多くのビットレートが割り当てられれば、MP 復号器を用いて左眼用シーケンスのみを観視する場合にも画質改善が期待できる。この場合、基本レイヤと拡張レイヤの和としては固定ビットレート伝送でありながら、基本レイヤでは符号化の難易度に適合した可変ビットレート伝送が行えることになる。

今後は、画質の主観評価試験を行い、左右の

SNR 値の平均値が最大の場合に立体視時の画質が最良となるとした仮定の妥当性を検討するとともに、MVP 方式の画質とビットレートの関係を明らかにしていく予定である。また、MVP 方式のように立体方式ではない復号器との両立性を持った符号化方式においては、単眼画像の画質も検討の対象とすべきである。このような観点からも検討を行っていく。

参考文献

- [1] "Information Technology - Generic Coding of Moving Pictures and Associated Audio: Video", ISO/IEC 13818-2 Amendment 3, WG11 N1366, Sept. 1996.
- [2] "Information Technology - Generic Coding of Moving Pictures and Associated Audio: Video", ITU-T H.262 (1995) | ISO/IEC 13818-2 (1996).
- [3] MPEG Test and Video Subgroup, "Results of MPEG-2 Multi-View Profile Verification Test", WG11 N1373, Sept. 1996.
- [4] W. J. Tam & L. B. Stelmach, "Perceived Image Quality of MPEG-2 Stereoscopic Sequences", proceedings of SPIE '97, 3016-34, Feb. 1997.