

立体ハイビジョンにおける視差補償符号化の検討

内藤 整 松本 修一
KDD 研究所

〒 365 埼玉県上福岡市大原 2-1-15
sei@avc.lab.kdd.co.jp

立体ハイビジョンの符号化に MPEG-2 マルチビュープロファイルを適用した場合、再生画質に関してチャンネル間のバランスが常に保たれるとはいえ、立体画像として評価した際に立体感が損なわれる可能性がある。

本稿では、各チャンネルの復号画像品質、および立体画像としての立体感を性能指標として、立体ハイビジョンの視差補償予測に基づく R チャンネルの符号化方式に関して検討を行った。その結果、比較的低レートで立体ハイビジョンの符号化を行う場合、マルチビュープロファイルでは、立体映像として十分な品質を得られない可能性があることを示す。

A study on the disparity compensated prediction coding method for 3-dimensional HDTV sequences

Sei NAITO Shuichi MATSUMOTO
KDD R&D Laboratories

2-1-15 Ohara Kamifukuoka-shi Saitama 356, Japan
sei@avc.lab.kdd.co.jp

When the MPEG-2 Multi-View Profile is applied as a coding method for 3-dimensional HDTV, the picture quality balance between 2 channels could always be kept.

In this paper, we investigate the coding method for R channel based on a disparity compensated prediction, evaluating the coding performance from the decoded picture quality of each channel and the obtained sense of perspective. As a result, we concluded that applying Multi-View Profile for coding 3-dimensional HDTV video sequences at comparatively low bitrate can not always supply enough picture quality as a 3-dimensional sequence.

1 はじめに

立体ハイビジョンは、左眼の視点、右眼の視点にそれぞれ対応するLチャンネル、Rチャンネルの2本のハイビジョン信号から成る。このようなステレオ動画像を符号化する場合、各チャンネル内での動き補償により行う時間方向の冗長さの圧縮に加えて、チャンネル間の視差補償により行う視差冗長さの圧縮が可能であり、各チャンネルを独立して符号化する場合よりも高い符号化効率が期待される。

視差補償は、他方のチャンネルにおける再生時刻が同一となるピクチャの復号画像を参照して予測を行うため、その適用対象はいずれか一方のチャンネルに限定される。以下では、図1.1に示すように、LR両チャンネルのうち、LチャンネルがLチャンネル内に閉じた動き補償のみにより符号化され、RチャンネルがRチャンネル内の動き補償、およびLチャンネルの再生画像を参照する視差補償により符号化される場合を想定する。図において、矢印は予測の向きを示す。

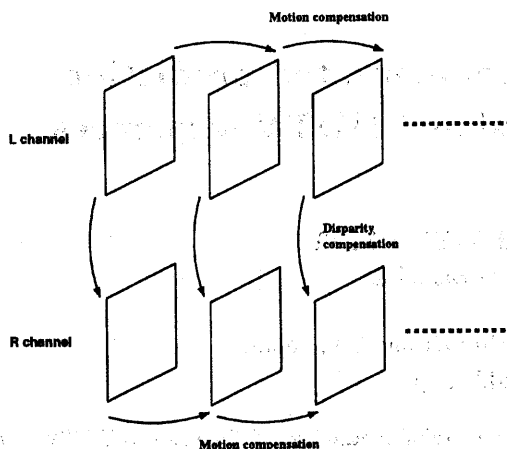


図 1.1: ステレオ動画像符号化

以上の考えに基づくステレオ動画像符号化の国際標準として、既にMPEG-2マルチビュープロファイルが規定されている [1]。本仕様に沿って符号化を行うことにより、各チャンネルを別々に符号化した場合(サイマルキャスト方式)と同等の立体再生画質が得られることが、主観評価により確認されている [2]。しかし、本方式では、ビットレートが各チャンネルに固定的に割り

当てられるので、チャンネル間の符号化効率の差が入力画像により大きく変動することを考慮すると、再生画質に関してチャンネル間のバランスが常に保たれるとはいえ、立体画像として評価した際に立体感が損なわれる可能性がある。

以上の問題点を踏まえ、本稿では、視差補償予測に基づくRチャンネルの符号化方式に関し、各チャンネルの復号画像品質、および立体画像としての立体感を性能指標として検討を行った。検討の対象としては、Rチャンネル符号化において、視差検出以外の部分で簡略化の可能性を検証する目的から、以下の2点を考えた。

- ・動き補償の適用方法
- ・予測誤差信号の符号化方法

2 Rチャンネル符号化方式

以下に示す5つのRチャンネル符号化方式を対象として、性能比較を行う。以下に各方式に関して簡単に説明する

方式 A

マクロブロック単位の予測を以下の3モード間で適応的に切替える。

- ・再生時刻が直前のRチャンネル復号画像を参照する動き補償予測
- ・再生時刻が同一のLチャンネル復号画像を参照する視差補償予測
- ・上記2枚の復号画像を参照する双方向予測

マルチビュープロファイルにおけるBピクチャの予測方法がこれに相当する。

動きベクトルの検出方法に関して簡略化の可能性を検証するため、方式Aをさらに2方式に分ける。動き補償で使用する動きベクトルの検出をブロックマッチングにより行うものを方式A-1、動きベクトルの検出を行わずにすべて0ベクトルとするものを方式A-2とする。予測の結果得られた予測誤差信号は、方式A-1、方式A-2ともにDCTにより符号化する。

方式 B

マクロブロック単位の予測を再生時刻が同一の L チャンネル復号画像を参照する視差補償予測に限定して行う。マルチビュープロファイルにおける P ピクチャの予測方法に含まれる。ただし、マルチビュープロファイルの P ピクチャは、視差補償予測、動き補償予測のいずれかをピクチャ単位で選択可能であるという点で本方式とは異なる。

予測誤差信号の符号化方法に関して簡略化の可能性を検証するため、方式 B をさらに 3 方式に分ける。予測誤差信号の符号化に関し、方式 A と同様 DCT により行うものを方式 B-1、固定量子化ステップサイズの DPCM により行うものを方式 B-2、誤差信号の符号化を行わず予測画像を再生画像とするものを方式 B-3 とする。方式 B-3 については、予測誤差信号を符号化しないため、視差検出の精度が最も大きく再生画像に影響する。このことを考慮して、検出単位を細かくとり、視差検出精度を向上させた場合の効果も調べるため、方式 B-3 については、検出単位を 16×16 のブロックとした場合、および 16×8 のブロックとした場合の 2 通りを考慮することとする。

表 2.1: R チャンネル符号化方式

方式	予測方式 (MB 単位)	動きベクトル の算出	予測誤差信号 の符号化
A-1	視差補償/	動き検出による すべて 0 と仮定	DCT
A-2	動き補償適応		
B-1	視差補償のみ	-	DCT
B-2			DPCM
B-3			符号化せず

3 シミュレーション結果

検討の対象として挙げた R チャンネル符号化方式に関し、符号化シミュレーションを行った。符号化器の構成を図 3.1 に示し、符号化シミュレーション条件を表 3.1 に示す。

チャンネル間の画質の差を極力抑える目的から、チャンネルごとの符号化レートは固定的には割り当てない

ものとする。このため、ピクチャごとに与えるターゲットビットレートの算出は、L, R 双方の GOP を統合して考え、GOP 内のトータル残ビット数、ピクチャタイプごとの残ピクチャ数、およびピクチャタイプごとの符号化難易度を TM5 の計算式に適用することで行う [3]。符号化は、L チャンネルの先頭ピクチャから始め、以後、R チャンネルピクチャと 1 枚ずつ交互に行うものとする。ここで、L チャンネル、R チャンネル双方の符号化部で仮想バッファを共有し、その占有量を参照してマクロブロックごとの量子化パラメータを更新するものとした。これにより、各チャンネルの符号化難易度が他方の量子化制御に反映されるため、符号化効率の違いによる再生画質の差もある程度吸収されるものと考えられる。

ただし、差分信号を固定量子化ステップサイズ DPCM により符号化する方式 B-2、および差分信号を符号化しない方式 B-3 に関しては、量子化制御を受けないため、再生画質差抑制の対象外とする。また、方式 B-2 における DPCM の固定量子化ステップサイズは、GOP 単位に割り当てられたビット数に不足が生じないという条件の下、設定可能な最小サイズにより決定した。

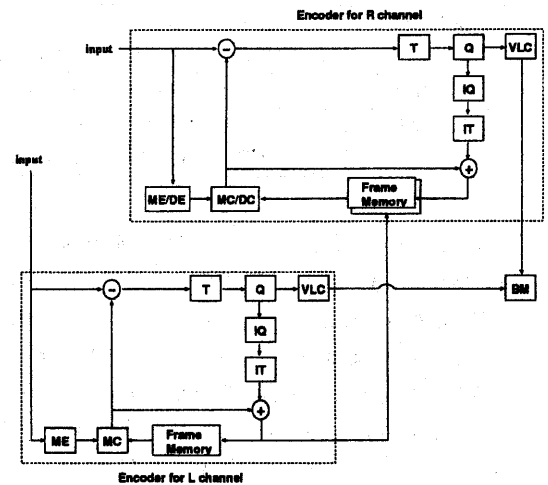


図 3.1: 符号化器の構成

入力となる立体ハイビジョンシーケンスとしては、以下の2種類を使用した。

● Soccer

サッカーのコーナーキックのシーン。カメラ操作として比較的高速なパニングを含み、画面中央部で複数の選手が絡む激しい動きを伴う。ゴールネットや背景となる観客席など、精細な部分も多く含む。

● Flowers

チューリップ畑の通路を女性がゆっくりと歩くシーンから成る。前方に位置する花や背景となる木の枝といった精細な部分を多く含むが、激しい動きを伴う部分はない。

表 3.1: シミュレーション条件

シーケンス	1920 画素×1036 ライン, 4:2:2
フレーム数	60 フレーム
符号化レート	40Mbit/sec
ピクチャ構造	フレームピクチャ
GOP 構造 (Lch)	N = 15, M = 1
動き補償	フレーム予測/フィールド予測適応
動き検出	ブロックマッチング法 ±127 画素×±31 ライン, 半画素精度
視差補償	フレーム予測/フィールド予測適応
視差検出	ブロックマッチング法 ±31 画素×±3 ライン, 半画素精度

表 3.2, 表 3.3 に各シーケンスに対する復号画像の SN 比の結果を示す。表中には、フレームごとにチャンネル間で確認された SN 比の差に関して、全フレームにおける平均値、最大値を合わせて示す。表 3.2, 表 3.3 の結果より、方式 B-1 (視差補償 + DCT) により R チャンネルの符号化を行った場合に、チャンネル間の SN 比の差が最も少なく抑えられていることが分かる。

表 3.4, 表 3.5 に各シーケンスに対するチャンネル別の発生符号量の結果を示す。表中の発生符号量は、60 フレーム (2sec) 分に相当し、各チャンネルの符号量の総和は方式によらずほぼ同一の値となっている。

復号画像を立体ハイビジョン用モニターに出力し、原画像との比較により主観画質評価を行った結果を以下

表 3.2: 復号画像の SN 比 (Soccer)

方式	復号画像 SN 比 dB		SN 比の差 dB	
	Lch	Rch	平均	最大
A-1	32.87	31.97	0.90	3.1
A-2	32.43	30.60	1.83	4.6
B-1	31.24	30.72	0.52	1.9
B-2	32.85	28.36	4.49	6.7
B-3(16×16)	34.56	26.92	7.64	10.3
B-3(16×8)	33.56	27.63	5.93	9.0

表 3.3: 復号画像の SN 比 (Flowers)

方式	復号画像 SN 比 dB		SN 比の差 dB	
	Lch	Rch	平均	最大
A-1	30.98	29.28	1.70	2.1
A-2	30.47	27.98	2.49	2.9
B-1	29.38	28.54	0.84	1.5
B-2	30.84	26.64	4.20	5.0
B-3(16×16)	32.86	25.12	7.74	8.2
B-3(16×8)	32.27	25.74	6.53	7.1

表 3.4: 発生符号量 (Soccer)

方式	発生符号量 ×10 ⁶	
	Lch	Rch
A-1	49.3	30.5
A-2	45.1	34.7
B-1	37.1	42.8
B-2	52.8	27.0
B-3(16×16)	68.3	11.6
B-3(16×8)	56.9	22.9

表 3.5: 発生符号量 (Flowers)

方式	発生符号量 $\times 10^6$	
	Lch	Rch
A-1	48.0	31.9
A-2	42.9	37.0
B-1	34.6	45.3
B-2	50.4	29.5
B-3(16 \times 16)	71.6	8.37
B-3(16 \times 8)	63.8	16.1

に示す。いずれの方式においても、画面全体として見た場合に明らかに立体感が失われているものはなかったが、一部の方式では、局所的なブロック歪みなど、無視できないレベルの劣化が確認された。

- 方式 A-1
Soccer については、ゴールネットの一部にちらつきがみられる。それ以外に関しては、特に劣化は確認されなかった。
- 方式 A-2
Soccer については、ゴールネットや観客席のフェンスにちらつきがみられ、Flowers については、背景の木や敷き詰められたチューリップに不自然な揺れがみられた。
- 方式 B-1
方式 A-2 の結果と同様、Soccer についてゴールネットや観客席のフェンスにちらつきがみられ、Flowers について、背景の木や敷き詰められたチューリップに不自然な揺れがみられた。方式 A-2 と比較して、その優劣を判断する明らかな違いは存在しないが、Soccer については、全体的に多少ぼやけた画像となっている。
- 方式 B-2
Soccer については、客席の一部にブロック歪みがみられ、Flowers については、動物体の境界がぼやけた画像となり、全体的に精細さが失われている。
- 方式 B-3
Soccer については、客席部分にブロック歪みが頻

発しており、Flowers については、動物体の境界にブロック歪みが多くみられた。このブロック歪みのうち、緩やかな動きを伴う動領域部分、花や背景部分の木の枝といった精細部分に存在するものは特に目立ち、視覚的に明らかに不自然であると感じるレベルのものである。視差ベクトルの検出ブロックサイズを 16×8 とすることで、ブロック歪みは感知されにくくなったが、それでもなお精細部分のブロック歪みは明らかに確認された。

4 考察

4.1 共有バッファ制御による効果

方式 A-1、方式 A-2、および方式 B-1 においては、共有バッファ占有量に応じた量子化制御が効いており、表 3.2、表 3.3 からも分かるように、符号化効率の違いによる再生画質の差が比較的小さな値に抑えられており、立体視を行う上でも支障がないレベルである。

これに対し、マルチビュープロファイルのように、各チャンネルへの固定ビットレート割当を前提とする場合、再生画質の差をあるレベルに抑えることが可能な最適比率がシーケンスごとに異なるのは、表 3.4、表 3.5 の結果からも明らかである。

以上の点から、ステレオ動画像へのビットレート割当は固定的に行うべきではなく、量子化制御を互いに連携して行うことにより、結果的に最適なビットレート比率を得ることが望ましいといえる。

4.2 動き補償の適用

表 4.1、表 4.2 に、R チャンネル符号化方式ごとの符号化モードの選択比率を示す。方式 A-1 に関して、動き補償予測のみのモードが選択される確率は 50% 近くまで達しているのに対し、視差補償予測のみのモードが選択される確率は 10% にも満たないことから、一般に視差補償予測単体の予測効果は動き補償予測単体に比べて低いといえる。さらに、方式 A-1 において視差補償予測と動き補償予測を双方向的に使用するモードの選択確率が 50% 近くまで達していることから、視差補償予測は、動き補償予測と併用可能な場合に十分な効果を発揮するものと考えられる。立体画像としての

主観評価の結果からも、方式 A-1 は明らかに方式 A-2、方式 B-1 を大きく上回っていることから、立体ハイビジョンを 40Mbps という比較的低レートで符号化する場合、R チャンネルに動き補償予測を併用することにより得られる効果は大きいといえる。

表 4.1: 符号化モードの選択比率 (Soccer)

方式	符号化モード %			
	Intra	視差予測	動き予測	視差/動き予測
A-1	0.5	8.7	47.9	42.4
A-2	2.9	63.6	22.5	10.5
B-1	9.8	83.5	—	—

表 4.2: 符号化モードの選択比率 (Flowers)

方式	符号化モード %			
	Intra	視差予測	動き予測	視差/動き予測
A-1	0.3	7.5	45.1	46.7
A-2	0.9	40.7	33.6	24.5
B-1	3.6	93.1	—	—

4.3 予測誤差信号の符号化

視差補償による予測誤差信号の符号化手法として、予測誤差信号を符号化しない場合、および量子化ステップサイズの粗い DPCM により符号化した場合には、DCT により符号化した場合に比べて L チャンネルの画質がやや向上するものの、R チャンネルの画質が極端に劣化し、これにより立体画像として評価した際にも明らかな劣化として確認される。

ステレオ動画像において、いずれか一方のチャンネルに発生したブロック歪みのような明らかな雑音は、立体画像として評価した場合にも検知されることから、R チャンネルにおける DCT による予測誤差信号の符号化に対し、本研究で検討したような大幅な簡略手法による置き換えの余地はないものといえる。

5 結論

本研究では、視差補償予測に基づく R チャンネルの符号化方式に関し、動き補償の適用方法、予測誤差信号の符号化方法を検討の対象とし、各チャンネルの復号画像品質、および立体画像としての立体感を性能指標として方式の評価を行った。

その結果、ステレオ動画像へのビットレート割当は、量子化制御を互いに連携して行うことにより、結果的に最適なビットレート比率を得ることができることを確認した。また、立体ハイビジョンを 40Mbps などの比較的低レートで符号化する場合、視差補償予測に基づく R チャンネルの符号化方式に対し、動き補償予測を併用すること、および予測誤差信号を DCT により符号化することは必須の要素技術であることを明らかにした。

一つの結論として、立体ハイビジョンを 40Mbps などの比較的低レートで符号化する場合、固定ビット割当を前提とするマルチビュープロファイルでは、R チャンネルの復号画質の大幅な劣化により、結果的に立体画像としての品質が保証されない可能性があることを示した。

謝辞

日頃御指導頂く国際電信電話 (株) 研究所の村上所長に感謝いたします。

参考文献

- [1] "Information Technology - Generic Coding of Moving Pictures and Associated Audio/Video", ISO/IEC 13818-2 Amendment 3(Sep.1996)
- [2] Test and Video Subgroup, "Results of MPEG-2 Multi-View Profile Verification Test", ISO/IEC JTC1/SC29/WG11 N1373(Sep.1996)
- [3] Test model Editing Committee, "Test Model 5", ISO/IEC JTC1/SC29/WG11 N0400(Apr.1993)