

画像特徴に応じた MPEG-4 ビデオ符号化方式

米山 暁夫 柳原 広昌 笹野 義二 中島 康之

株式会社 KDDI 研究所
埼玉県上福岡市大原 2 - 1 - 1 5
Tel: 049-278-7423
E-mail: yoneyama@kddilabs.jp

本稿ではシーンの変化やシーンの特徴に適応した MPEG-4 ビデオ符号化アルゴリズムを検討したので報告する。本提案では入力画像数枚をプリロードし、そのプリロードした画像のマクロブロック (MB) 毎の画面内アクティビティおよびこのアクティビティを利用した動き特徴の解析により適切な I-, P-, B-VOP の配置を決定することを目的としている。通常は入力画像のシーンや動きの特徴を取得するためには演算量の増大が避けられないが、本方式では一般的なエンコーダで利用される情報を活用するため、演算量の増大を抑えた符号化が可能となる。

MPEG-4, 動き適応, シーンチェンジ, シーン適応

SceneMotion Adaptive MPEG-4 Coding algorithm

Akio Yoneyama Hiromasa Yanagihara Yoshiji Sasano Yasuyuki Nakajima

KDDI R&D Laboratories Inc.
2-1-15 Ohara Kamifukuoka-shi Saitama, 356-8502
Tel: 049-278-7423
E-mail: yoneyama@kddilabs.jp

In this paper, we propose an encoding algorithm for MPEG-4 video encoding. In our proposal, predetermined number of input pictures is preloaded. Then picture characteristics are analyzed and appropriate VOP types (I, P, B) are determined before encoding. Since the processing speed of the proposed algorithm is almost equal to that of the conventional MPEG-4 encoder without adaptive VOP types control, fast encoding can be realized with better picture quality than that of without the control of VOP types.

MPEG, Motion Adaptive, Scene Change, Scene Adaptive

1. はじめに

MPEG-1 や MPEG-2 は Video-CD, DVD, デジタル衛星 / 地上波テレビ放送など様々なアプリケーションでの動画の伝送や蓄積に幅広く利用されている。また、インターネットやモバイルでの利用を想定した様々なサービス、アプリケーションでは MPEG-4 ビデオ符号化方式が普及しつつある。MPEG-4 では、時間解像度（フレームレート）、画像空間的解像度、符号化ビットレートについて非常に広いレンジに対応している。従って、符号化制御についてもこのような広いレンジで最適な符号化効率を実現することが望ましいといえる。

また、符号化遅延は符号化構造により変化するが、許容される遅延量はアプリケーションの用途に依存する。例えば双方向でのテレビ会議システムでは符号化遅延を最小化する必要があるのに対して、片方向通信アプリケーションであるライブ放送型システムの場合には、双方向での通信と比較して符号化遅延の許容度が大きくなる。このような場合には、予め入力画像を数枚プリロードすることにより入力画像特徴をあらかじめ取得する方法も提案されている。文献[4]では最大予測フレーム間隔に相当する画像の先読みより予測フレーム間隔を適応的に設定する方式を提案している。

更に VOD 配信のような場合には許容遅延量はさらに増加できる。また、この場合にはオフライン符号化も利用可能なため、マルチパスエンコーディングのように、画質および符号化効率を向上させるための様々な処理の適用が可能となる。マルチパスエンコーディングでは、第一符号化パスにおいて入力シーケンス全体の解析を行い、ビット割当量を決定し、第二符号化パスでこのビット割り当てに従って符号化を行う[3,12]。

本稿では、リアルタイムでのエンコードを目的とし、許容される遅延量に応じてプリロード枚数を設定し、そのプリロードした画像の特徴を解析することによる“シーンアダプティブな VOP タイプ制御”と“適切な使用ビット量予測”アルゴリズムをベースとした MPEG-4 符号化方式を提案する。

本方式では“P-VOP 間隔: M ”、“I-VOP 間隔: N ”、“VOP あたりの目標ビット割当量: R ”の3つのパラメータを適応的に制御する。

2. VOP タイプの割り当て

MPEG 規格では符号化ビットストリームシンタックスについてのみ規定されているため、画質の向上を目的とした多くのレート制御方式

が提案されている[1,2,12]。また、符号化画質の更なる向上を目的として、ピクチャ符号化タイプや VOP 符号化タイプに着目した方式も提案されている。文献 [5,6,7]ではシーンの特徴に応じた適応的な I ピクチャの挿入による符号化方式が提案されている。

同様に MPEG-4 符号化においても、I-, P-, B-VOP を自由に配置することが可能である。Fig.1 にシーンチェンジを含む画像の符号化における、固定 I-VOP 挿入間隔と可変 I-VOP 挿入間隔を用いた場合における VOP 符号化タイプと典型的な符号量を示す。Fig.1 上側の例のように I-VOP がシーンチェンジの位置に関係なく挿入される方式では、I-VOP 挿入直後にシーンチェンジがあるような場合には、実質的には I-VOP 符号化相当の画像が必要以上に近接することとなり、全体としての符号化効率の低下を招く。一方、可変 I-VOP 間隔で符号化した場合には I-VOP の出現をシーンチェンジ位置に同期させることによって符号化効率の向上が可能となる。

次に Fig.2 にシーンの動き特徴が変化する画像の符号化における固定 P-VOP 間隔と可変 P-VOP 間隔での符号化時の VOP 符号化タイプと典型的な符号量を示す。一般的なエンコーダではさまざまなシーンで比較的良好な画質が得られる符号化設定として Fig.2 の上側の例のように、シーケンス全体に渡って固定的に P-VOP 間隔を 0.1 秒程度(30fps での符号化を想定した場合には $M=3$) で符号化することが多い。しかし、動きの激しい画像についてはピクチャ間隔が広がるにつれて急激にフレーム間相関が低下するため、 $M < 3$ とする方が、 $M=3$ の場合よりも符号化効率を改善することができる [8]。従って、動きの特徴が変化するシーンの場合は、P-VOP 間隔を適応的に変更することで符号化効率を改善させることが可能となる。このような

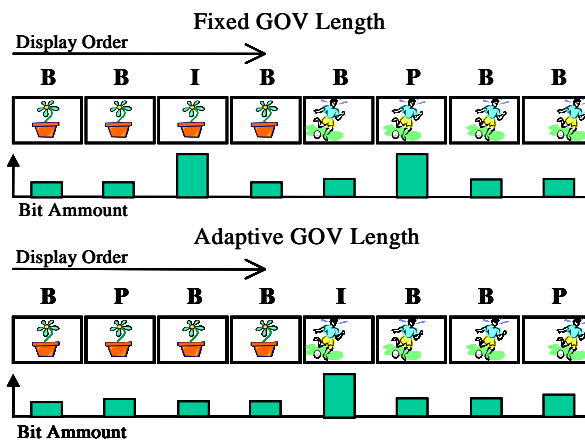


Fig.1 Fixed / Adaptive GOV length encoding

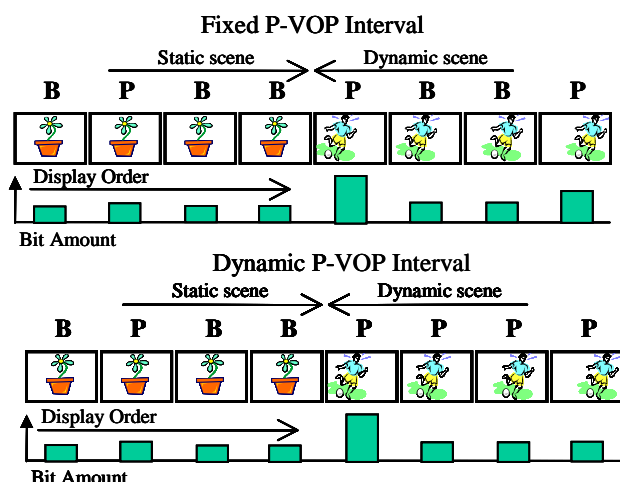


Fig.2 Fixed / Dynamic P-VOP interval encoding

可変 GOP 符号化方式として、たとえば文献[4]では、あらかじめ P ピクチャの最大予測間隔に相当する複数枚の画像を先読みし、その先読みした画像の特徴から P ピクチャ間隔 M を決定しながら符号化を行っている。しかしこの場合、I ピクチャ符号化直後にシーンチェンジがあるような場合の符号化効率の低下を避けることができない。筆者らはこれまで MPEG-1 および MPEG-2 のビデオ符号化において、シーンチェンジに適應させた可変 GOP 構造における固定ビットレート符号化方式[10]および可変ビットレート符号化方式[13]を提案した。これらの提案においては、プリロードした画像情報から 1GOP 内の各ピクチャタイプを決定し、符号化を行うが、その際の GOP 内の P ピクチャの間隔は一定値としていた。本稿では文献[13]における方式を更に拡張し、MPEG-4 符号化においてシーンに応じて動的に GOV 内の P-VOP 予測間隔を任意に切り替える方式を検討する。

以下に提案方式の符号化フローを示す。

ステップ 1 : 入力画像のプリロード

- 1) あらかじめ最大 I-VOP 間隔 N_{max} を設定する。
- 2) プリロード枚数を N_{pl} とし、 N_{pl} 枚の入力画像をプリロードする。

ステップ 2 : フレーム内符号化複雑度解析

- 3) 画面内アクティビティ平面を生成し、その平面からテクスチャおよびフレーム内符号化における符号化複雑度の指標を得る。

ステップ 3 : シーンチェンジ検出と I-VOP 挿入位置の決定

- 4) 3)よりプリロードした画像内でのシーンチェンジ検出を行う。
- 5) 検出されたシーンチェンジの位置により、

I-VOP の挿入位置を決定する。

ステップ 4 : 動き特徴解析

- 6) 次に符号化する画像を Pi_{Cn} とし、その画像を含めた 3 枚(1/10 秒相当)の画像の両端である Pi_{Cn} と Pi_{Cn+2} から生成された空間アクティビティ平面を利用し、簡易動き探索を行うことで動き特徴の解析を行う。

ステップ 5 : P-VOP 間隔の決定

- 7) 6)の結果より、 Pi_{Cn} 、 Pi_{Cn+1} 、 Pi_{Cn+2} の 3 枚の入力画像に対して P, P, P での符号化を行うか、B, B, P での符号化を行うかの判定を行う。

ステップ 6 : 目標ビット量の割当て

- 8) 目標ビットレート、バッファ占有量、VOP タイプ、過去の VOP の符号化結果を元に、対象 VOP の目標ビット量を決定する。

ステップ 7 : VOP 符号化

- 9) 目標ビット量と使用ビット量の割合から各 MB に対する量子化パラメータ値を調整し、 Pi_{Cn} 、 Pi_{Cn+1} 、 Pi_{Cn+2} の 3 枚の符号化を行う。
- 10) 次の 3 枚の入力画像を読み込みステップ 2 へと戻る。

以下に各ステップの詳細を説明する。

2.1 入力画像のプリロード

まずシーケンスのエンコードにおける最大 I-VOP 間隔 N_{max} を設定する。この値はコンテンツのランダムアクセス性、または符号化コンテンツの編集再利用の観点から最適値を決定する。

エンコードの開始時には、あらかじめ N_{pl} 枚の入力画像を先読みする。この N_{pl} の値はエンコードを開始する前に画像の特徴を解析するために必要な枚数を指定することになるため、リアルタイムエンコードの場合には符号化時の許容される遅延により制限を受けることになる。

また、この N_{pl} を大きく設定することにより、先の画像の変化をあらかじめ把握することができるため、シーンの変化に対する画質の変動を抑えることが可能である。

2.2 フレーム内複雑度の解析

画面内の MB 単位のアクティビティ値 (空間アクティビティ値) は、画像の特徴をよく表す指標として MPEG-2 TM5 符号化制御[9]等にも利用されている。本稿では画面内符号化における符号化複雑度の指標に上記空間アクティビティ値を利用する。また、可変 I-VOP 間隔および可変 P-VOP 間隔での符号化を行うためには、入

力画像からのシーンチェンジ検出とテクスチャ・動き特徴の検出が必須となるが、どちらも画素レベルでの処理を行った場合には演算量の増大を招く。

そこで、本提案ではシーンチェンジ検出および動き特徴解析にも、文献[13]同様に画素平面ではなく上記の空間アクティビティ値で構成される空間アクティビティ平面を活用する。

2.3 シーンチェンジ検出と I-VOP 挿入位置の決定

プリロードした N_{pl} 枚のフレームから空間アクティビティ平面を生成した後、I-VOP 符号化する画像を決定するために以下の2つの処理を行う。まずシーンチェンジの検出を行い、次にI-VOP 符号化する画像の特定を行う。シーンチェンジの検出は、適切な符号化制御を行う上で以下の観点から必要となる。

- 1) シーンチェンジにより発生ビット量の増加が起こるため、ビット配分を変更する必要がある。
- 2) シーンチェンジ前後での画像特徴に沿って相関がなくなるため、符号化時のパラメータはリセットされる必要がある。

このシーンチェンジ検出において高い精度を実現するために様々な方式[5,6,7]が提案されているが、本提案では上述の空間アクティビティ平面を利用して検出を行った。隣接する画像間で同一位置のMBの持つ空間アクティビティの絶対差分値を求め、1画面内の平均値があらかじめ指定された閾値よりも大きい場合には、その画像間がシーンチェンジであると判断する。

次にI-VOPの挿入位置を決定する。I-VOP 間隔の均一化を目的とし、以下の条件により決定する。

$$\begin{aligned} & \text{If } (n_{sc} > N_{max}) \text{ then } N = \max(n_{cur}, n_{sc}/2) \\ & \text{Else } N = n_{sc} \end{aligned} \quad (1)$$

ここで、現在の符号化対象 VOP 以前の最近 I-VOP からの距離を n_{cur} とし、検出されたシーンチェンジ位置までの距離を n_{sc} とする。また顕著なシーンの変化がない場合には、I-VOP 挿入間隔は N_{max} に設定する。

2.4 動き特徴の解析

最適な P-VOP 間隔を決定する最も効果的な方法としては、組み合わせが可能な全ての P-VOP 間隔を用いて実際に符号化を行い、その中から最も符号化効率の高い設定を選択すること

であるが、これは膨大な探索処理を必要とするため現実的とはいえない。そこで本提案ではステップ2で得られた空間アクティビティ平面上での簡易な動き予測誤差を算出する。この簡易な動き予測誤差の算出に用いる画像の時間的間隔は、 $M=3$ のときの P-VOP 予測間隔である $1/10$ 秒とする。空間アクティビティ平面上で1サンプル毎に探索範囲を水平垂直 ± 1 サンプル以内（原画像上では水平垂直 ± 16 ピクセルに相当）として絶対誤差和(SAD)の計算を行う。本方式における動き探索では精度の高い動き情報そのものを得ることはできない。しかしこの動き探索における予測誤差 E_{ave} が入力画像の変化の激しさを表す特徴値として利用することができる[10]。以下に 352×240 サイズの画像における動き特徴解析と本提案で用いたアクティビティ平面上での特徴解析における1ピクチャあたりの加算処理回数の比較結果を示す。

- i)探索範囲 ± 16 画素全探索 :
of addition / picture : 1.8×10^8
- ii)探索範囲 ± 16 画素 TSS :
of addition / picture : 5.5×10^6
- iii) MB 数 22×15 アクティビティ平面 ± 1 全探索 :
of addition / picture : 3.0×10^3

以上から分かるとおり、本提案で利用する簡易探索(iii)は画素領域での高速な探索方式(ii)と比較しても $1/1000$ 以下と非常に演算量が少なくなっている。

2.5 P-VOP 間隔の決定

文献[10]において、筆者らはPピクチャ間隔 M を決定するためにアクティビティ A_{ave} と簡易動き探索により得られた予測誤差 E_{ave} を用いて以下のようにモデル化した。

$$M = M_{max} - b \cdot (E_{ave} / A_{ave}) + c A_{ave} \quad (2)$$

本提案では3枚の画像に対しての符号化方法を P,P,P とするか B,B,P とするかを切り替えるため、式(2)において求めた値を閾値(Thr_M)処理により M を切り替える。

以上により決定した I-VOP 挿入位置と、P-VOP 間隔から符号化対象画像の VOP タイプを決定することができる。

2.6 目標ビット量の割当

目標ビット量の決定は、同一シーンにおける I-,P-,B-VOP の符号化複雑度情報と、平均量子化パラメータ、およびプリロード時に算出している空間アクティビティ情報からのフィードバ

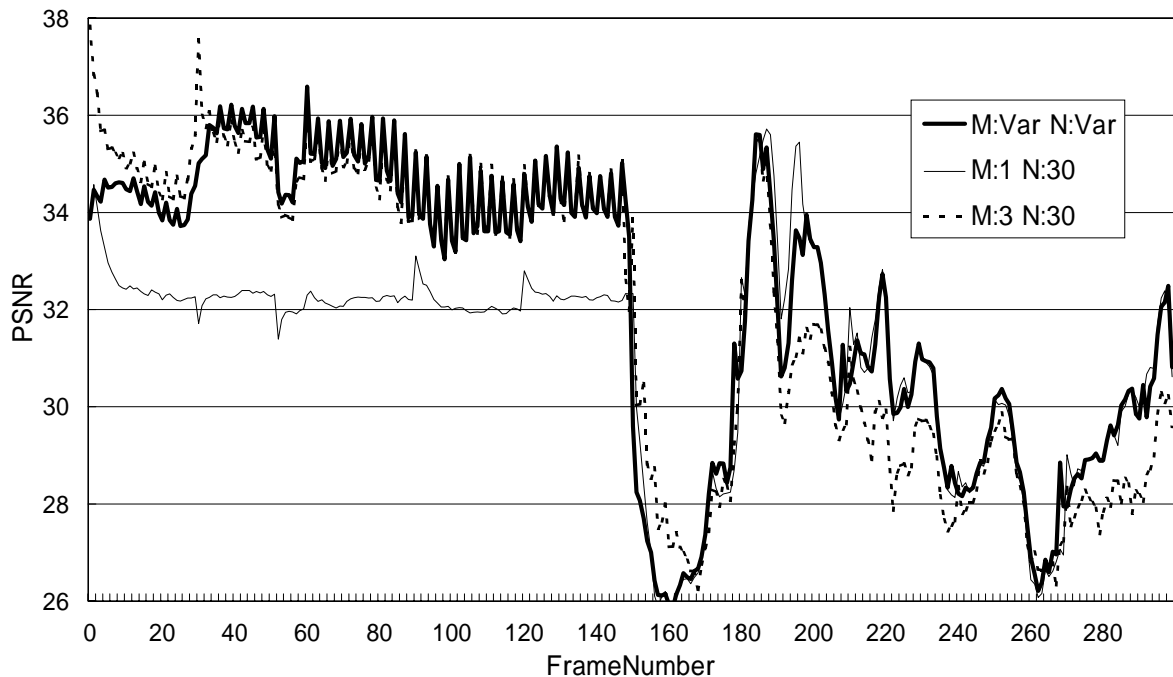


Fig.3 PSNR results for various M, N Setting

ック制御を行う。ベースとなる方式には MPEG-2 TM5 を用いる。また、プリロードしている画像内でシーンチェンジを検出した場合には、シーンチェンジ以前での使用目標ビット量を低減させ、シーンチェンジ直後のビット消費に対応する。

シーンチェンジ後の画像に対しては、シーンチェンジ前の画像特徴を示すパラメータの適用は適切とはいえない。そのため I-,P-,B-VOP のビット配分割合を決定する重み付けパラメータ (w_I, w_P, w_B) については、入力画像の特徴解析で求めた各画像の空間アクティビティ値の平均値 A_{ave} と、そのアクティビティ平面を利用して前述のとおり求めた簡易動き補償予測誤差量 E_{ave} との比率を利用し、この比率 E_{ave} / A_{ave} が大きいほど I-,P-,B-VOP での所要ビット量の割合が均一に近くなることから、 $w_P = 1$ を基準として $w_I = I * A_{ave} / E_{ave}$ 、 $w_B = B * E_{ave} / A_{ave}$ と設定した。

3. シミュレーション

提案アルゴリズムにおけるパラメータの設定値を Table.1 に示す。

Table.1: Parameters in the algorithms

Npl	$Nmax$	$Mmax$	b	c	Thr_M	l	B
20	30	4	20	0.02	1.4	0.25	5.0

Table.1 の中で、パラメータ M_{max}, b, c は文献 [13] と同様の値を設定し、それ以外のパラメータについては事前の長時間のテレビ番組を入力画像とした予備実験を行い統計的に決定した。本方式を検証するために MPEG 評価用画像を利用し、DIVA : 150 frames から Football : 150 frames にシーンチェンジする画像を生成し、(size:320x240, fr:29.97fps Bitrate: 384kbps) で符号化を行った。Fig.3 は提案方式である M : 可変, N : 可変での符号化を行った際の PSNR を固定 I-, P-VOP 間隔である $M : 1, N : 30$ 、 $M : 3, N : 30$ での符号化結果と比較したものである。 $M : 1, N : 30$ と $M : 3, N : 30$ の PSNR 結果から、動きの少ない前半(DIVA)では $M : 3$ が有効であり、動きの激しい後半(Football)では $M : 1$ が有効であることがわかる。また、提案方式においては P-VOP 間隔の判定により前半では $M : 3$ をとり、後半においては 260 フレーム付近以外では $M : 1$ をとっていることを確認した。この動的な切り替えにより動きの特徴に関わらず PSNR を維持していることがわかる。

Table.2 は各方式における PSNR 結果を前半および後半の平均値に分けて表したものである。

Table.2: PSNR statistics of three methods

PSNR	M :Var, N :Var	M :1, N :30	M :3, N :30
1 st Half	34.6dB	32.3dB	34.7dB
2 nd Half	29.8dB	29.9dB	29.1dB

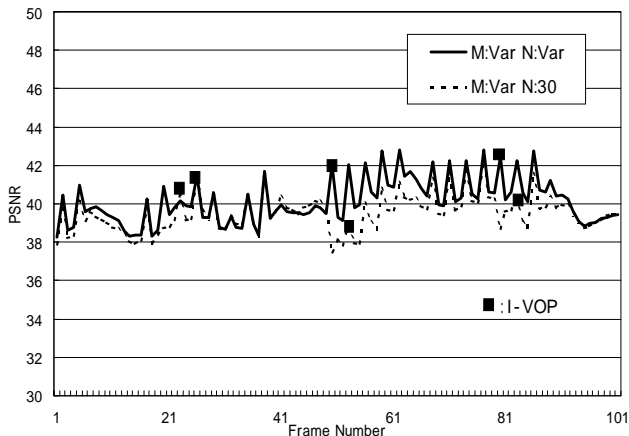


Fig.4 PSNR result of variable/fixed N

Table.2 より、提案方法は前半後半それぞれにおいて0.1dBの差があるが、これはシーンチェンジ前におけるビット配分の制限が影響していると考えられる。次に可変 I-VOP 間隔と固定 I-VOP 間隔との比較結果を Fig.4 に示す。入力画像にテレビ画像からキャプチャしたシーンの切替りも多く、また映像の動きの変化も激しいサッカーのシーンを利用した。符号化設定は320x240pixel, 29.97fps, 384kbpsとした。Fig.4はその符号化結果から特徴的な部分を抜き出したものである。このなかで、50フレーム目にシーンチェンジを含んでいる。図中の黒四角印は符号化タイプがI-VOPであることを示している。どちらの方式も P-VOP 間隔は動的に設定されているため、シーンの内容により3枚を最小単位として切替っているが、I-VOP 挿入間隔が固定となっている場合には図中の点線に示されるとおり、シーンチェンジ直後(50フレーム目直後)においてPSNRが大きく低下している。一方I-VOP 挿入間隔を動的に設定している場合には、シーンチェンジに関わらずPSNRを維持していることがわかる。また、その他のテレビ番組等からの入力画像を利用した検証においても、シーンの切替りの多い画像の場合には提案方式の有効性が顕著であることを確認した。

また符号化に伴う処理時間については固定 I-,P-VOP 間隔での符号化と比較して、3%程度の増加量に抑えられていることを確認した。

5. まとめ

本稿では、入力画像特徴に応じて I-VOP の挿入間隔および P-VOP 間隔をを適応的に変更することで符号化効率を改善する MPEG-4 符号化方式を検討した。これらのパラメータ、各 VOP に対する目標符号化ビット量の決定には、MPEG-2 TM5 ベースの符号化制御時に通常利

用される MB 単位の輝度アクティビティ値を指標とした画面内符号化複雑度情報、輝度アクティビティ値で生成される平面上での簡易動き予測誤差情報による動き特徴解析により行う。また実験により提案方式では平均 PSNR を改善していることを確認した。

謝辞

日頃よりご指導いただき、(株)KDDI 研究所 浅見徹代表取締役所長、松本修一取締役、和田執行役員に感謝いたします。

参考文献

- [1] L. Luo, C. Zou, Z. He, I. Shirakawa, "An Adaptive Quantization Algorithm for MPEG-2 Video Coding", Proc. ICASSP'98, Vol.5, pp.2841-2844, 1998.
- [2] N. Zhang, C. Wu, "Augmenting STM5 Bit Rate Control Scheme for Improving Coding Quality of MPEG2 Video Encoders", Proc. ISCAS'97, Vol.2, pp.1297-1300, 1997.
- [3] L. Teixeira, H. Ribeiro, "Analysis of a Two Step MPEG Video System", Proc. ICIP'97, Vol.1, pp.350-353, 1997.
- [4] Y. Yokoyama, Y. Ooi, "A Scene-Adaptive One-Pass Variable Bit Rate Video Coding Method for Storage Media", ICIP'99, Vol.3, pp.827-831 Oct., 1999.
- [5] Y. Yu, J. Zhou, Y. Wang, "A Fast Effective Scene Change Detection and Adaptive Rate Control Algorithm", Proc. ICIP'98, ICIP 1998.
- [6] M. Lee, S. Kwon, J. Kim, "A Scene Adaptive Bitrate Control Method in MPEG Video Coding", SPIE VCIP, Vol.3024, pp.1406-1416, 1998.
- [7] S. Bozoki, R.L. Langendijk, "Scene Adaptive Rate Control in a Distributed Parallel MPEG Video Encoder", ICIP'97, Vol.2, pp.780-783, 1997.
- [8] B.G. Haskell, A. Puri, and A.N. Netravali, "Digital Video : An Introduction to MPEG2", Chapman & Hall, 1996.
- [9] "MPEG-2 Test Model 5", ISO/IEC JTC1/SC29/WG11/ 93-400, Apr., 1993.
- [10] A. Yoneyama, *et al.* "MPEG Encoding Algorithm with Scene Adaptive Dynamic GOP Structure", MMSP'99, pp.297-302, Sep. 1999.
- [11] Kondo, Fukuda, "A Real-Time Variable Bit Rate MPEG2 Video Coding Method for Digital Storage Media", IEEE Trans. Consumer Electronics, Vol.43, No.3, pp.537-543, Aug. 1997.
- [12] W. Ding, B. Liu, "Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling", IEEE Trans. Circuits and Systems for Video Technology, Vol.6, No.1, pp.12-20 Feb. 1996.
- [13] 米山他, "シーンアダプティブな可変 GOP 構造を利用した1パスVBR MPEG符号化方式", 映像学技報 CE, Vol.25, No.61, pp.13-18 Sep. 2001.