

動画像の特徴抽出に基づく
高能率符号化の研究
High Efficiency Coding
Based on Video Feature Extraction

2005年2月

秦泉寺久美

目次

第1章	序論	1
1.1	研究の背景	1
1.2	従来技術とアプローチ	2
1.2.1	フレーム符号化型アプローチ	4
1.2.2	オブジェクト符号化型アプローチ	5
1.2.3	階層符号化型アプローチ	7
1.3	高能率符号化の観点からの従来技術の問題点	8
1.4	本論文の構成	9
第2章	本研究の目的とアプローチ	10
2.1	本研究の目的	10
2.2	スプライト符号化を用いたアプローチ—第3章～第6章—	10
2.3	FGS符号化を用いたアプローチ—第7章—	11
第3章	カメラ操作の自動抽出法の提案と多層スプライト生成	13
3.1	はじめに	13
3.2	カメラパラメータと平行移動動きベクトルの関係	15
3.2.1	カメラモデル	15
3.2.2	スプライトモデル	20
3.3	カメラパラメータ算出アルゴリズム	20
3.3.1	拡大縮小パラメータの算出	21
3.3.2	回転パラメータの算出	22
3.3.3	移動パラメータの算出	23
3.4	スプライト生成アルゴリズム	24
3.4.1	カメラの位置が固定の場合	24
3.4.2	カメラの位置の移動がある場合	26
3.5	カメラパラメータ算出実験	27
3.5.1	実験	27
3.5.2	結果並びに考察	29
3.6	スプライト生成実験	30
3.6.1	実験	31

3.6.2	結果並びに考察	34
3.7	本章のまとめ	34
第4章	グローバルモーション算出法の改良と各符号化方式への適用	36
4.1	はじめに	36
4.2	GM 算出アルゴリズムの改良	36
4.2.1	緩い輝度勾配マクロブロックの除去	38
4.2.2	拡大縮小パラメータの算出	38
4.2.3	回転パラメータの算出	38
4.2.4	移動パラメータの算出	39
4.3	改良アルゴリズムによる GM 算出実験	40
4.3.1	改良アルゴリズムの評価法	42
4.3.2	実験結果および考察	45
4.4	画像符号化への適用	45
4.4.1	符号化方法	46
4.4.2	実験結果および考察	46
4.5	本章のまとめ	51
第5章	2層ビデオオブジェクト自動生成とスプライト符号化への適用	52
5.1	はじめに	52
5.2	2層ビデオオブジェクト自動生成アルゴリズム	54
5.2.1	GM 算出	55
5.2.2	背景スプライト生成	55
5.2.3	前景オブジェクトの抽出	56
5.3	MPEG-4 符号化実験	59
5.3.1	前景比率と符号量の関係	61
5.3.2	フレームレートと符号量の関係	61
5.3.3	同一ビットレート符号化時のフレームレートと画質における評価	63
5.4	本章のまとめ	65
第6章	スプライトを適応的に利用した MPEG-4 映像符号化システム	67
6.1	はじめに	67
6.2	スプライトを適応的に用いた符号化システムの提案	68
6.2.1	スプライトショット符号化	68
6.2.2	スプライトショット符号化に適したショットの条件	70
6.2.3	ショット単位での処理の流れ	70
6.2.4	提案符号化システムの概略	71
6.3	スプライトショット抽出アルゴリズム	72
6.3.1	グローバルモーションによる2次判定	72

6.3.2	前景比率による3次モード判定	73
6.4	2層ビデオオブジェクトレート制御法	74
6.4.1	符号化実験	75
6.5	提案符号化システムの評価	75
6.5.1	評価画像	75
6.6	本章のまとめ	79
第7章	DCT係数を用いた主観品質に寄与する特徴の推定とスケーラブル符号化への適用	81
7.1	はじめに	81
7.2	MPEG-4 スケーラブル符号化	82
7.2.1	MPEG-4 スケーラブル符号化器構成	82
7.2.2	FGSにおけるSelective Enhancement	83
7.3	差分DCT係数と画像特徴	83
7.3.1	差分DCT係数分布の特徴	84
7.3.2	平坦領域とそれ以外の領域の区別	84
7.4	主観画質に寄与する画像特徴抽出アルゴリズム	86
7.4.1	主観画質を考慮した画像の領域の分類	86
7.4.2	画像特徴抽出アルゴリズム	87
7.4.3	画像特徴抽出実験	89
7.5	スケーラブル符号化実験	90
7.5.1	ビットプレーンシフトにおける基本方針	90
7.5.2	符号化実験	91
7.6	本章のまとめ	96
第8章	結論	97
	謝辞	100
	参考文献	101
	業績	107
	図一覧	111
	表一覧	114

第1章 序論

1.1 研究の背景

1990年代の前半，インターネットや携帯電話が爆発的普及を遂げ，その後，DSLの普及に伴い一気にネットワークの広帯域化が進んだ．そして，これらの環境下で画像の通信を行おうという動きが1990年代の後期に出てきた．総務省による「平成15年度通信利用動向調査の結果」[1]によれば，2004年1月現在，インタ-ネット加入者の総数が7700万人を超えると報告されている．そのアクセス回線の内訳は，自宅の場合，xDSLや光によるブロードバンドアクセスが半数をしめ，残りの半数をISDN，電話回線などの狭帯域回線が占める．また，個人のインターネット利用においてはパソコンからの利用が約6000万人，携帯電話，PHS，情報端末からの利用者が約4500万人を占める．また，「インターネット白書2004」によれば，ブロードバンドアクセスの内訳は，60%以上がxDSLによるものであると報告されている．

これらの調査報告から，確実に光アクセスを中心としたメガビット級のインターネットが普及しつつあるが，依然として大多数を占めるのはxDSLや携帯電話などの中低レート網からのアクセスであると言える．今後も，全てのアクセスラインが光ネットワークに移行するには時間を要し，また，光化が進んでも，帯域をシェアするIPネットワークでは，加入者が増えれば一人当たりが利用できる実効帯域が必然的に狭まると予想される．これらの通信環境下にてテレビ品質(VGA:640×480画素，30 frame per sec.)の画像を伝送することは困難である．蓄積・放送用に標準化されたMPEG-2を用いた場合，実時間で通信するためには最低でも数Mbpsの伝送容量を必要とする．画像を1Mbps以下の符号量まで効率よく圧縮する必要がある．

この符号量をターゲットとする画像圧縮符号化技術として，従来のテレビ電話に用いられたH.261，その後継であるH.263，蓄積用に標準化されたMPEG-1という方式が1990年代初頭から順次標準化されて来た．これらの圧縮符号化技術は，動き補償(Motion Compensation: MC)による予測誤差を離散コサイン変換(Discrete Cosine Transform: DCT)し，変換係数を量子化した値に可変長符号化(Variable Length Coding: VLC)を施すという基本的なパラダイムに基づく．このパラダイムにおけるMCは，垂直および水平方向への幾何的な動きをモデル化(平行移動モデル)したものであり，このモデルに当てはまらない動きを持つ画像は圧縮効率が必然的に低下する．

一方で，1999年末から標準化されてきた画像符号化方式であるMPEG-4[3]は，高圧縮率と共に機能が重視された最初の画像符号化標準である．MC+DCTという符号化パラダイムを超えた，新しい符号化ツールや機能が提案された．符号化ツールとしては，スプライト，グローバル動き補償，オブジェクト符号化などが新たに加えられた．その結果，従来は矩形のフレームを

単位として符号化されていたが、任意の形状を持つビデオオブジェクト単位に符号化することが可能になった。新しい機能としては、無線伝送環境での伝送特性を考慮し、リバーシブル可変長符号化やビデオパケット構造の導入など、誤り耐性が強化された。また、一つのビットストリーム（符号化データ）から複数のビットレート（伝送速度）のビットストリームを取り出すことができるスケーラブル符号化 [4] も標準化された。これらの新しい符号化パラダイムによって、従来符号化方式にはない劇的な圧縮効率や機能を実現することが可能になった。

しかし、新しい符号化ツールが全ての画像に対して従来符号化パラダイムである MC+DCT 以上に効果を発揮するわけではない。これらの符号化ツールを効果的に利用するには、符号化を行う前段にて、画像の特徴（性質、構造など）を明らかにし、その特徴に合致した符号化ツールを選択する必要がある。テレビ電話のような画像、非常に高速なカメラ操作がある画像、フォローショット画像（被写体をカメラで追尾した画像）、線画や均等色領域から構成されるセルアニメーション画像はそれぞれ異なる特徴を持つ。画像の特徴に応じた符号化ツールを適用することによって、1Mbps 程度の伝送帯域でも主観的、客観的に高い品質を実現できると考えられる。

本論文は、以上を研究の背景とし、インターネットやモバイル通信に適用するために、画像を効率よく符号化することを目的とする。本論文では、特に、新しい符号化ツールのうち、オブジェクト符号化とスケーラブル符号化に着目する。オブジェクト符号化は画像を空間的に複数の領域に分割し、各々を独立に符号化する方式である。これらの領域をビデオオブジェクトと呼ぶ。各々のビデオオブジェクトを合成することで元の画像が得られる。スケーラブル符号化は画像を画素値の振幅（大きさ）の方向に分割する。この分割した画像をレイヤと呼び、各レイヤ毎に符号化を行う。各レイヤを合成することで元の画像が得られる。双方とも、画像をビデオオブジェクトやレイヤに分割して符号化する。画像を符号化モデルに合致するように分割することで、符号化効率の向上が期待できる。本論文は、画像から圧縮に適した特徴を抽出する方法とそれを実際に高圧縮符号化に適用する方法について研究し、その成果をまとめたものである。

1.2 従来技術とアプローチ

画像の圧縮効率の向上や符号化データから意味をもつものを取り出すことを目的として、1990年代より従来の符号化パラダイム（MC+DCT）を超えたアプローチが提案されている。齊藤 [5] らは、「画像内容」を符号化パラメータとすることを念頭に、符号化モデルを Low Level, Mid Level, High Level の三つに分類した。Low Level 符号化は、信号波形をそのまま用いるもので、従来の MC+DCT 符号化がこれに相当する。Mid Level 符号化は画像における幾何学的（パラメトリック）な動き特徴をモデル化したもので、カメラ操作などの幾何学的な動きをする領域を符号化情報から取り出せる。1995 年前後をピークに Mid Level 符号化に相当する研究が様々なところでなされている。High Level 符号化は任意の動きをする個々の物体や、照明の変化など画像が CG 並にパラメータ化されたものである。符号化はなんらかの意味をもつ塊毎に行われる。齊藤らの分類は、画像中の「意味のあるもの」を符号化に利用することと、符号化データから「意味のあるもの」を取り出すことに主眼が置かれている。よって、画像の圧縮効率の優劣について議論したものではない。これに対し、本論文では、画像の圧縮効率向上の観点から、符号化アーキ

表 1.1: 画像特徴を用いた符号化のアプローチ

符号化アーキテクチャ	符号化ツール	利用する画像特徴
フレーム符号化型	イントラ・インター MB 判定 I,P,B ピクチャ切替 GMC, DS	予測誤差 大局的な動き
オブジェクト符号化型	オブジェクト符号化, スプライト グローバルモーション	背景の大局的な動き 前景の動領域
階層符号化型	FGS	ROI, 視覚的に重要な部分 (エッジ, グラデーション)

テクチャの特徴によって、符号化方式をフレーム符号化型、オブジェクト符号化型、階層符号化型の三つのアプローチに分類した。その様子を表 1.1 に示す。

フレーム符号化型アプローチは、フレーム毎に順次符号化を行う従来型の符号化アーキテクチャを基本とする。符号化ツールとしては、フレーム内で符号化が閉じているフレーム内符号化と、隣接フレームとの相関を利用する予測符号化が該当する。予測符号化には主に MC+DCT が用いられる。フレーム符号化型アプローチに有効に働く画像特徴は「予測誤差」である。予測誤差の大小によってイントラ・インター符号化モードを切り替えたり、I, P, B などのピクチャタイプの切り替えることによって圧縮率の向上が可能となる。一般的にはその部分の符号量がより小さくなる符号化ツールが選択される。この方法は、領域分割をしたり、画像を階層化する必要はない。古くは、マクロブロックのイントラ・インターモード判定、ピクチャタイプの適応的切替がある。また、近年提案された新しい符号化ツールでは、グローバル動き補償 (Global Motion Compensation: GMC), ダイナミックスプライト (Dynamic Sprite: DS) 等のパラメトリックな動き補償法がこの方法に相当する。パラメトリックな動きそのものがカメラモーション等の画像の大局的な動きである場合があり、パラメトリックな動きが画像特徴として利用される。

オブジェクト符号化型アプローチは、画像を時間・空間方向に領域分割することによって構造化を行い、対応する領域毎に最も適した符号化方法を採用する方法である。これらの一連の領域はビデオオブジェクトと呼ばれる。符号化は領域毎に閉じている。このアプローチでは、任意形状のオブジェクト符号化、スプライトなどの符号化ツールが使用される。また、背景部分のカメラ操作を反映した大局的な動きや、「人物」「動物体」など視覚的に意味をもった背景以外の動領域が画像特徴として利用される。本論文では便宜上「前景動領域」と呼ぶことにする。このアプローチでは、単独の領域から元の画像の全体像を再現することができない。

階層符号化型アプローチは、もとの画像の概略を表す画像 (基本レイヤ) と詳細情報を表す画像 (拡張レイヤ) からなるように画像を階層 (レイヤ) 化する。基本レイヤの情報に拡張レイヤの情報を付加することで画像の解像度を高めたり (空間スケーラビリティ), フレームレート (1 秒間に符号化するフレームの枚数) を増したり (時間スケーラビリティ), 品質を向上させたり (SNR スケーラビリティ) する。時間・空間的な領域分割は必要としないため、基本レイヤの

みで元の画像の概略を再現することが可能である。MPEG-2における階層符号化，MPEG-4のFine Granrality Scalable (FGS) 符号化 [4]，JPEG2000等がこのアプローチに相当する。ここで用いる画像特徴としては，信号処理レベルのものから，視覚的に意味のあるROI(Region Of Interest)まで幅広い定義が可能である。

次に，三つのアプローチにおける先駆的な研究を紹介する。

1.2.1 フレーム符号化型アプローチ

古くは，マクロブロックのイントライナーモード判定，ピクチャタイプの適応的切替がこれに相当する。前者においては，ほとんどの標準的な符号化方式に採用されている。また，後者においては，Iピクチャを適応的に挿入するための画像特徴であるカット点を抽出する方法等 [6]が提案されている。また，PピクチャとBピクチャの適応的切替法としては，高速な動きのある場合は，Pピクチャを適用し，画像が静止している場合や非常にゆっくりとした動きのある場合はBピクチャを適用する方法 [7]が提案されている。この場合の画像特徴は画像の大局的な動きである。この大局的な動きの大小によって適用するピクチャタイプを判定している。

前述の古典的な手法に加え，近年提案されてきた符号化ツールと従来符号化ツールを適応的に切り替える方法が提案されている。その符号化ツールの中で注目されるのが，GMCやDSである。これらの符号化ツールで利用する画像特徴はグローバルモーションと呼ばれるものである。グローバルモーションは，領域の変化を一組の幾何変換パラメータで表すもので，「パラメトリックな動き(変換)」と表現されることもある。

パラメトリックな変換の種類としては，以下のものが挙げられる。

- 平行移動モデル (rigid model, 2パラメータ)
- アフィンモデル (affine model, 6パラメータ)
- ヘルマートモデル (hermart model, 4パラメータ, アフィンモデルの特殊な場合)
- 射影モデル (planer model, 8パラメータ)

平行移動モデル，アフィンモデル，射影モデルとも，着目するマクロブロックと，過去のフレームの領域に関して，それぞれ平行移動，アフィン変換，射影変換を施して得られる予測マクロブロックとの差分を最小とする変換パラメータを動きベクトルとする。グローバルモーションは撮像時のカメラ操作を反映する場合が多く，人間の直感との親和性が高い。

グローバルモーションを抽出する研究には，90年代の初めにBergen[8]らによって行われた。その後，Irani[10]，Wang[11][12]，上倉 [13][14]，Smolic[15][16]らが続く。これらの研究は，大きく2つのアプローチがなされてきた。一つはあらかじめ算出された局所的(ローカル)な動きベクトルをクラスタリングし，最大クラスタのセントロイド(重心)をグローバルモーションとするボトムアップ的な手法 [11][12][13]である。もう一つは画面全体を変形して，直接的にグローバルモーションを算出するトップダウン的な手法 [10][9]である。前者は，局所的な動きの精度がグローバルモーションの精度に大きく寄与する。よって，輝度勾配のなだらかな画像などの局所的

な動きが正確に算出できない画像に対しては、確実にカメラ操作を反映したグローバルモーションを算出することが困難であった。また、最大クラスがインライヤ（カメラモーションのみを反映したデータ）を代表しているとは限らなかった。一方、後者は、原画像と予測画像の差分の評価関数を最小化する変換パラメータがグローバルモーションとして抽出される。一般的に、評価関数には、差分画像の全体の画素の2乗和の平均値（MSE: Mean Square Error）や絶対差分平均値（MAD: Mean Absolute Difference）採用される。この方法は局所的な動きベクトルの抽出精度に拠らない。しかし、アウトライヤ（カメラモーションを反映しない領域）をあらかじめ削除して予測画像を評価する必要がある。アウトライヤを削除しない場合は、アウトライヤも含めた最適なグローバルモーションが算出される。よって、カメラ操作のみを反映する領域において、評価関数が最小になるとは限らない。しかしながら、アウトライヤを自動的に抽出することは、それ自体動き領域を抽出する別の課題となるので、非常に困難な問題であった。

パラメトリックな動きを用いた符号化の例としては、パラメトリックな動きを一つのフレームに対して一つだけ用いる Irani[10][9] や上倉 [13]、如澤 [17][18][19] の方法がある。この方法は、後に MPEG-4 の GMC 符号化として標準ツールに採用された。

1.2.2 オブジェクト符号化型アプローチ

オブジェクト符号化は、あらかじめ画像を領域分割し、その領域毎に符号化を行う。これらの時空間的な領域をビデオオブジェクトと呼ぶ。

MPEG-4 におけるビデオオブジェクトは任意の形状を持つことができる。ビデオオブジェクトはテクスチャ符号化と形状符号化から成り、これらは独立に符号化される。テクスチャは、矩形になるように境界の画素値を用いて埋められ（パディング）、テクスチャ情報は従来方法の MC+DCT 方式で符号化される。各々のビデオオブジェクトの特徴によって、例えば、注視している部分 (ROI: Region of Interest) が明らかであれば、背景部分の符号量を減らして、ROI の符号量を多くして主観画質をあげることが可能である。ビデオオブジェクトの生成は標準化の対象外である。ビデオオブジェクトの構成例としては、ビデオオブジェクトは個々の物体を反映したオブジェクト（便宜上、前景オブジェクトと呼ぶ）と個々の物体ではなく、カメラ操作のみを反映した領域を表現するオブジェクト（便宜上、背景オブジェクトと呼ぶ）に分類する方法がある。本論文の第2章以降では、この前景オブジェクトと背景オブジェクトを研究対象にしている。一般的なビデオオブジェクトの抽出方法に関しては、以下の3つに分類できる。

- 完全に手動で行う方式
- 半自動で行う方式
- 完全自動で行う方式

世の中にソフトウェアプロダクトとして実際に普及しているビデオオブジェクト抽出手法はほとんどが完全手動方式である。一方、半自動方式は、Kass らが SNAKES[20] といわれる初期輪郭を手動で設定することによって、その後の一連の画像におけるオブジェクト領域を推定するという研究が発端となり、以降、いろいろなところでその改良方式が発表されるに至る。

その中で、特に注目されるのは、ウォーターシェッド（分水嶺）アルゴリズムを基本とする Salembier[21][22][23][24] の研究である。これは、マーカ（初期輪郭）を利用し、ダムの中に水を浸していき、周囲にエッジがあるとそこで領域の拡張を止めるものである。境田 [25] や酒井 [26]、Choi[27] らもウォーターシェッドアルゴリズムを応用した領域抽出法を提案している。これらの半自動的に抽出する方法は完全手動で抽出する方式に比べて、オーサリングにかかるコストが格段に減少した。しかし、意とする領域を完全に抽出するには改善の余地がある。また、世の中に存在する膨大なコンテンツをターゲットにするには依然として処理コスト大きい。

一方、完全自動処理に関しても、動作を保証するための拘束条件が常に存在した。その多くはカメラが固定されていて常に同じ背景の元で撮像されているという条件である。境田 [25] や Neri[28][29]、Garth[30]、Aach[31]、Mech[32]、Vincent[33] らも同様の拘束条件を必要とする。

一方で、「背景」部分もひとつのオブジェクトとしてとらえることができる。一般的な前景オブジェクトが任意形状を持ち、ランダムな動きをするのに対し、背景オブジェクトはカメラ操作を反映したパラメトリックな動きをもつことが多い。背景オブジェクトを前景オブジェクトと同様に、任意形状のビデオオブジェクトとして扱うことも可能である。しかし、パラメトリックな動きを用いた方が圧縮効率は高い。以下に、パラメトリックな動きを用いた場合の背景オブジェクトの表現方法を説明する。パラメトリックな動きによって、画像を変形しながら張り合わせると、一枚のモザイク画像（ビデオモザイク：Video Mosaic）を生成することができる。モザイク画像はパノラマ画像、スプライト（Sprite）などと呼ばれる。本論文では、これらを統一して、スプライトと呼ぶことにする。スプライトは、元来はCG用語であった。ここで、CGにおけるスプライトの使い方を2通り示す。例えば蝶の羽を表現するときに、小さな断片（スプライト）を用意しておき、これを幾何変換させることで、蝶の羽がばたばた動く様を作成するという例である。もうひとつの例は、2次元テレビゲーム等に見られる巨大な背景（スプライト）としての利用である。これは、スプライトの必要な部分を幾何変換して切り出して背景画像にする、というものである。どちらの方法も、スプライトをあらかじめ用意しておき、スプライトを幾何変換して用いる。大きなスプライトを用いる場合は、幾何変換した後に、必要な部分（画像でいうフレーム）だけ切り出すという処理が加わる。これを画像符号化に適用するためには、基本となるスプライトと、時間毎に必要な部分を幾何変換して切り出すためのグローバルモーションが必須である。

スプライトの生成は Tedosio[34] らの研究に端をなす。彼女らは、ビデオシーケンス中の2次元動きベクトルの集中に着目し、これをトラッキングすることでパノラマが生成されることを述べている。スプライトを符号化に適用した例としては、Lee[35] らの研究がある。数百 kbps の符号量にて、従来方法に比較して劇的な主観品質を達成した。彼らの提案をもとにスプライト符号化が MPEG-4 符号化のツールに採用された。Wang, Adelson[11][12] らは一つのフレームに対して複数のパラメトリックな動きを用いて、複数のスプライトを生成し、“Mobile calendar”等の標準画像に適用し、その圧縮効果を示唆している。また、上田ら [36] はセッティング表現という、スプライトの3次元的な配置法を考案し、画像圧縮に適用するための検討を行った。しかしながら、これらの従来技術 [9][10][11][35] は、あらかじめ抽出されたパラメトリックな動きを持つ領域のみをターゲットとしており、任意の自然画像への適用については言及されていなかった。

その他のビデオモザイク生成においては，阿久津 [37][38][39]，谷口 [40]，星野 [41]，竹内 [42][43] がそれぞれのアプリケーションに特化した方式の提案を行い，良好な結果を得ている．パノラマ画像は人間の直感との親和性が高いことを利用し，これを画像のインデックスとする研究 [40][38][41] が行われている．ビデオモザイクングを利用してビデオカメラの解像度以上の高精細な静止画像の生成法 [41][42][43] の提案がなされている．

1.2.3 階層符号化型アプローチ

階層符号化の歴史は長く，動画像符号化方式が標準化される前の，JBIG における静止画符号化にも階層符号化の概念が取り入れられている．

一方，動画像においては，MPEG-2 では階層符号化 (= スケーラビリティ) の概念が標準画像符号化方式の中に初めて取り入れられた．MPEG-2 には時間スケーラビリティ，空間スケーラビリティ，SNR スケーラビリティの概念が取り入れられている．符号化方式は基本レイヤと拡張レイヤの 2 階層方式である．この三種類のスケーラビリティのうち，画像品質の向上に寄与するものは SNR スケーラビリティである．拡張レイヤは，基本レイヤと原画像の残差信号を符号化する．しかし，MPEG-2 の枠組みでは，拡張レイヤから一部のビットストリームを取り出すことは不可能である

MPEG-4 においては，基本レイヤに MPEG-4 Simple/Advanced Simple Profile を，拡張レイヤには Fine Granularity Scalable (FGS) Profile [4] を規定し，SNR スケーラビリティを実現している．FGS 符号化は一つのビットストリームから複数のビットレートのビットストリームを取り出せることが MPEG-2 と異なる．

スケーラビリティはインターネット等の帯域の変動，パケットロス，遅延，ジッタ等が存在する通信環境にて画像を伝送する場合に，ビットストリームの一部を捨てても画像が途切れることなく受信・再生できる機能を実現する．一方で，基本レイヤにて符号量が不十分な部分を拡張レイヤにて補ってトータルな画像品質を向上させることも可能である．基本レイヤにて SNR の悪化する部分を補填するのが通常の拡張レイヤの方法であるが，SNR の大小が必ずしも主観品質に関係があるわけではない．この点から，画像のどのような部分を拡張レイヤによって補填するかという課題は依然として残る．

上記の課題に対して，田邊 [44] らは，MPEG-2 符号化において ROI 領域を自動的に抽出し，階層的に符号化することによって，低レートでの主観品質の向上を試みている．一方，中嶋 [45] らは FGS 符号化において，マクロブロックの優先順位を変更できる Selective Enhancement (S.E.) に着目し，視覚的に妨害が目立つエッジや高輝度領域などを画像特徴として抽出し，その領域を優先的に符号化する方法を考案している．しかし，田邊らの研究は，ROI の定義と実際の主観品質の評価が必ずしも一致しないという問題点があった．また，中嶋らの研究は画像信号から画像特徴を直接抽出しているため，基本レイヤの画質にかかわらず均一な特徴しか得られないという問題点があった．

1.3 高能率符号化の観点からの従来技術の問題点

本論文では、画像特徴を抽出し、その結果を適応的に画像符号化に取り入れ、符号化効率の向上を目的とする。本節にて、画像の圧縮効率向上に観点から、各アプローチにおける従来技術の問題点についてまとめて述べる。

三つのアプローチのうち、フレーム毎符号化型アプローチについては、マクロブロックの符号化モードの選択やピクチャタイプの選択等の主要な方式は既に実用化レベルにあり、検討の余地は少ない。また、パラメトリックな画像特徴を用いた符号化法である GMC や DS に関しては、次に述べる理由によって検討対象から除外した。画像の動きが平行移動モデルである場合、動きベクトルの符号化データ全体に占める割合が大きい、極低ビットレートを除いては MC+DCT を符号化パラダイムとする従来符号化法と符号化効率が変わらないと推定できる。また、平行移動モデル以外に当てはまる大局的な動き（カメラのズームや回転など）が少ないと考えられる。適応しても動きモデル以外の要因による影響（例えば、ズーム画像においては、GMC や DS は効果的と考えられるが、極端にズームする画像では、カメラの動きによるボケが生じる）が加わると考えられる。ゆえに、パラメトリックな画像特徴を用いたとしても、符号化効率の劇的な向上は望めない。以上の理由により、本論文では、フレーム毎符号化型のアプローチは検討対象外とした。

オブジェクト符号化型のアプローチにおいて、ビデオオブジェクトを何らかの方法にて抽出した後、ビデオオブジェクト単独で符号化する方法は試みられているが、画像特徴抽出方法の提案、もしくはビデオオブジェクト生成への利用にとどまり、汎用的動画像符号化への適用はなされていなかった。オブジェクト符号化を汎用符号化として実現するには、汎用的な特徴抽出アルゴリズムの検討、ビデオショット単位での符号化ツールの適応性の判定、複数ビデオオブジェクトのレート制御法など、各課題を解決する必要がある。本論文では、スプライトの劇的な高圧縮性に着目し、オブジェクト符号化型アプローチを採用し、スプライトを適応的に符号化へ取り込むことによる汎用符号化システムの構築を目指す。

階層符号化型のアプローチは、画像に付加情報を用いることで機能性を向上させるというのが、従来技術の主目的であった。また、画像に機能性を持たせることで、分割損（符号化効率の劣化）が生じることは衆知であり、機能性と符号化効率はトレードオフの関係にあるとされてきた。両者を生かす手法、すなわち、情報を付加することで符号化効率そのものを向上させようという試みはなされていなかった。階層符号化においては、インターネットやモバイル環境での画像伝送が今後も主流になると予想され、これらの通信環境に適した符号化方式のあり方として、スケーラブルな機能は着目するに値する。特に、ビットストリームが低レートでカットされたとしても、主観的な画質を劣化させないスケーラブルビットストリームの構築は望まれるところである。本論文では、今後の需要を鑑み、階層符号型のアプローチからも画像特徴を利用した符号化効率の向上を目指す。

1.4 本論文の構成

本論文ではオブジェクト符号化型のアプローチ（第3章～6章）ならびに階層符号化型のアプローチ（第7章）から、画像特徴の抽出とそれを適応的に利用した符号化方式の検討を行う。本論文の構成と概要を以下に示す。

第2章にて、本研究の目的とアプローチについて述べる。

第3章にて、動画像からの大局的な動き（グローバルモーション）抽出法を提案し、これを多層スプライトの生成に適用する方法を述べる。7種類のカメラ操作とグローバルモーション、局所的な動きの間の関係を定式化する。

第4章では、背景画像のスプライト符号化への適用を目的とし、スプライト生成に適したグローバルモーション算出法の提案を行う。また、画像中の背景部分（背景画像）に対して、実際にグローバルモーションを用いた数種類の符号化方式（GMC, DS, SS）とグローバルモーションを用いない従来符号化方式との符号量の比較を行う。

第5章では、前景領域抽出法を提案し、画像から前景と背景を自動的に切り出し、それぞれをオブジェクト符号化、スプライト符号化を施す、2層ビデオオブジェクトモデルを提案する。2層ビデオオブジェクトモデルによる符号化方式を「スプライトモード」と呼ぶ。スプライトモードが従来符号化法（ノーマルモード）に比して、劇的な圧縮効率をもたらす条件を符号化シミュレーションより導き出す。

第6章では、スプライトが効果的であるショットを自動的に抽出する方法を取り入れ、スプライトを用いたショットと従来符号化法を適応的に切り替えることによる汎用画像符号化システムを構築する。

第7章では、FGS符号化に用いられる、原画像と基本レイヤの符号化画像の差分をDCT変換した係数（差分DCT係数）から、主観品質に影響を与える画像特徴を推定し、その結果を符号化に適用する。ビットストリームが低レートでカットされても視覚的な劣化が少ないビットストリーム構成法を提案する。

第8章では、本論文の結論を述べる。

第2章 本研究の目的とアプローチ

2.1 本研究の目的

本研究は、インターネットやモバイル通信に適用するために、画像を効率よく符号化することを目的とする。本研究では、新しい符号化ツールのうち、オブジェクト符号化とスケーラブル符号化に着目する。前者は従来符号化技術に比較して、劇的な圧縮効率をもたらす可能性がある。また、後者は、機能性と符号化効率のトレードオフが問題となったが、機能性と符号化効率の両者を生かせる可能性がある。双方とも、画像をビデオオブジェクトやレイヤに分割して符号化する。本論文は、画像から圧縮に適した特徴を抽出する方法とそれを高圧縮符号化に適用する方法について研究し、その成果をまとめたものである。以下にスプライト符号化を用いたオブジェクト符号化型アプローチおよびFGS符号化を用いた階層符号化型アプローチによる課題解決のアウトラインを示す。

2.2 スプライト符号化を用いたアプローチ—第3章～第6章—

次世代動画像圧縮の標準であるMPEG-4では、新たにVOP (Video Object Plane) という概念を採り入れ、背景や物体等のビデオオブジェクト毎に符号化することが可能である。特に動画像中の各々のフレームに共通な領域の動きが一組の平面の動きモデルで表される時、スプライトと呼ばれる平面オブジェクトを生成することが可能である。すなわち、背景画像に代表されるような動画像中の複数フレームにわたる領域を一枚の静止画であるスプライトから再構成できる。スプライトを使用すると符号化効率の大幅な改善が期待できるため、スプライト生成技術に関する期待は大きい。本論文の第3章にて、ビデオクリップを撮影した時のカメラモーションに着目し、(1)7種類のカメラモーションと平行移動動きベクトルの関係を定式化し、(2)カメラモーションを反映した大局的な動きを算出するアルゴリズムを提案、(3)大局的な動きを用いてスプライトを自動生成するアルゴリズムを提案した。また、複数の実画像を用いて実験を行なった結果、撮影時のカメラモーションを良好に抽出し、これを用いてスプライトを自動生成することを示す。

第4章にて、グローバルモーション算出およびスプライト生成の精度の向上を目指す。スタティックスプライトを用いた符号化では、スタティックスプライトの生成精度が符号化効率、SNR、主観評価に大きく影響する。動画像からスタティックスプライトを生成するにはアウトライヤの影響のないカメラモーションだけを反映したグローバルモーションを算出する必要がある。本論文では、局所的動きベクトルとその微分値の特徴空間での分布の特性を利用し、グローバルモーションの複数候補による検定手法を用いた、アウトライヤの影響の少ないグローバルモーション算出

法を提案する。また、提案手法を用いて算出されたグローバルモーションを MPEG-4VMVer.11 に適用して、(1)LME 符号化、(2)GME 符号化、(3) ダイナミックスプライト符号化、(4) スタティックスプライト符号化を行い、それぞれの符号量と画品質を評価した。その結果、スタティックスプライト符号化は LME 符号化、GME 符号化、ダイナミックスプライト符号化に比べて高圧縮率を達成できることを示す。

第5章にて、ビデオクリップを背景スプライトと前景オブジェクトに自動的に分離する方法を提案し、ビデオクリップ全体での符号量の評価を行う。生成した前景、背景のビデオオブジェクトをそれぞれ MPEG-4 オブジェクト符号化ならびにスプライト符号化に適用（スプライトモード）して、スプライトを用いない MPEG-4 Simple profile の符号化法（ノーマルモード）と比較検討を行う。特に、フレームレートや前景オブジェクトの割合を変化させ、符号量に与える影響を調査する。その結果、フレームレートに関係無く、前景比率が画面全体の 10-15 % 程度である場合に MPEG-4 通常符号化方法の $1/4 \sim 1/2$ 程度の符号量で同程度の画質を実現できることを確認する。また超低レート (128Kbps, 64Kbps) においてフレームレート、客観画質の評価を行い、同程度の SNR で倍以上のフレームレートを達成できることを示す。

第6章にて、さらに一般化し、あらゆる画像に対して、スプライト符号化を適応的に利用して圧縮効率を高める手法を提案する。スプライト符号化は、効果的な場合は、従来符号化法に比べて劇的な符号量の削減をもたらすが、スプライトがどのような画像にも効果があるとは限らないということも知られており、適応的な利用が望まれていた。第6章では、スプライトを適応的に符号化へとりこみ、MPEG-4 Main Profile 準拠の符号化システムとして開発する。本符号化システムは、スプライトが効果があると推定されるショットを自動的に抽出するスプライトショット抽出法を取り入れ、符号化品質の向上を図った。また、ストリーミング配信への適用も考慮し、低遅延スプライトに適用する2層ビデオオブジェクトレート制御法を新たに考案した。これにより、コンスタントビットレートの符号化データの生成を可能とする。また、カメラの動きのあるスポーツ映像を用いて、本提案符号化システムの評価を行い、その優位性と適応条件について考察する。

2.3 FGS 符号化を用いたアプローチ—第7章—

一つのビットストリーム（符号化データ）から複数のレートのビットストリームを取り出せるような符号化をスケラブル符号化という。また、このような利用を One Source Multi Use という。本論文の第7章では、一つのビットストリームから低いレートでビットストリームが切り出された場合でも、主観的な品質の劣化を抑制するスケラブルビットストリームの構築を目的とする。スケラブル符号化として、基本レイヤに MPEG-4 Advanced Simple Profile、拡張レイヤに MPEG-4 Fine Granularity Scalable (FGS) Profile を用い、拡張レイヤの符号化の優先順位を変更する手法である Selective Enhancement(SE) を利用して、品質の向上を図る。FGS 符号化では、基本レイヤの符号化画像と原画像の差分信号を DCT 変換した係数（差分 DCT 係数）を符号化する。本研究では、差分 DCT 係数と画像の特徴の関係を明らかにし、画像の主観品質に寄与するブロックノイズおよびモスキートノイズが出現するブロックを差分 DCT 係数および

ローカルデコード画像から推定する画像特徴抽出法を提案した。また、本画像特徴抽出法と符号化される差分 DCT 係数の特徴から、SE における 2 つの自由度である (1) どの領域に適用するか、(2) どのくらい優先するかを決定するビットプレーンシフト法を提案する。また、本提案手法を自然画像およびセルアニメーション画像に適用し、スケーラブル符号化実験を行い、PSNR とノイズの現れ方から画像の品質の評価を行う。その結果、従来の FGS 符号化を用いた場合に比べて、主観品質を向上させることを実証する。

第3章 カメラ操作の自動抽出法の提案と多層スプライト生成

3.1 はじめに

動画像の圧縮符号化において、連続するフレームの背景画像を一枚の静止画として符号化することで符号化効率を向上させる方法が検討されている [9][10][11][12][36][35]。動画像符号化の次世代標準方式 MPEG-4[3] においても、従来の符号化に無い新しい技術としてスプライト¹を使用した符号化方式が提案されている [3][35]。動画像符号化におけるスプライトとは、ビデオクリップ（シーンチェンジを含まない動画像）の各々のフレームに共通の大局的な領域をパノラマ画像的に表現したもので、モザイクイメージ [9][10]、レイヤ [11][12] 等の名で呼ばれている。これらの共通の領域は平面の動きモデルとして表せられる。個々の動きを含まない背景画像や、CGなどで制作したタイトルロゴなどがスプライトとして表せられる。一枚のスプライト（静止画）で動画中の複数フレームにわたるオブジェクトの動きを表現できるので、スプライトを使用すると同等の画質において既存符号化手法に比べて大幅な圧縮が可能であることが報告されており [35]、スプライト生成技術への期待は大きい。

本章では、カメラモーションのみを反映した背景画像をスプライトの対象とし、スプライト生成法について新たな技術の提案を行なう。スプライトは以下のように生成される。最初に、ビデオクリップの複数フレームに共通な領域の動き（大局的動き、グローバルモーション）を抽出する。次に、抽出されたグローバルモーションを用いて各フレームを幾何変換し、任意の基準座標にパノラマ的に順次貼り付けることで生成される。グローバルモーション抽出技術はスプライト生成の中核を占めている。

このグローバルモーションの算出法に着目すると、スプライト生成に関する従来技術においては大きく2つの手法が考案されている。ひとつは、隣接フレーム間で画像全体を変形して大局的な動きを算出するもので、トップダウン的なアプローチと呼ばれている [9][10]。もうひとつのアプローチとして、局所的な動きを特徴空間にてクラスタリングし、各々のクラスタにおける動きの代表をグローバルモーションとする方法がある [11][12][36]。これはボトムアップ的なアプローチと呼ばれている。ここで、カメラ操作と大局的な動きの関係の観点から両アプローチの優位点と問題点を述べる。

カメラモーションには静止している状態を除くと一般にパン（カメラの横振り）、チルト（同、立て振り）、ズーム（焦点距離の移動）、ロテート（レンズの光軸を中心とした回転）、トラック（カメラの位置の水平方向への移動）、プーム（同、垂直方向への移動）、ドリー（同、前後の移

¹スプライトはもともと CG（Computer Graphics）の分野で用いられていた概念である。

動)の7種類がある。これらは表3.1に示すように撮像位置の移動の有無によって2つのカテゴリに、また、画像の変形に着目した場合、拡大縮小、平行移動(水平・垂直)、回転の4つのカテゴリに分類できる。ここで、カメラの撮像位置の移動を伴わないパン、チルト、ズーム、ロテートの4種類の組合せで表せられるカメラ操作を用いて撮像された画像は、視差やオクルージョン(領域の見えかくれ)を伴わないので、背景画像を一枚のスプライトとして表現ができる。一方で、カメラの撮像位置の移動を伴うトラック、ブーム、ドリーを含む画像は実世界の構造に応じて視差やオクルージョンを生じる。このような画像からスプライトを生成するには、複数フレーム間で領域を対応づける必要がある。完全自動で領域を対応づけることは困難であり、この問題を解決すべく初期領域を手動で抽出し、半自動的に領域分割を行なう方法[11]、あるいは領域分割処理はすべて手動で行なうといった方法[35][36]が考案されていた。カメラの撮像位置が変化する場合、1枚の平面そのものを撮像していない限り、複数の領域が抽出されるのが一般的である。

トップダウン的アプローチはビデオクリップ全体から一組の大局的動きを算出する方法で、もともと一枚のスプライト(平面)しか仮定していない。よってカメラの撮像位置が変化する場合カメラ操作には適用不可能であった。また、アウトライヤ(カメラモーション以外の動きを反映している外れデータ)の影響の観点から、トップダウン的アプローチはフレーム全体を変形して元の画像と変形した画像の各画素における誤差の和を最小にする変換パラメータを最適値とするために、アウトライヤの影響を受けやすいという問題があった。

一方、ボトムアップ的アプローチは複数の領域の個々の動きからそれぞれスプライトを生成することが可能だが、前述のように実世界を反映した領域分割処理を完全自動で行なうことは困難である。よって、完全自動なスプライト生成は不可能であった。これは大量の映像素材への適用を考慮すると望ましくない。さらにボトムアップ的アプローチにおいては、局所的にアフィンや透視投影モデルによって動きを算出して、その個々の動きを特徴空間でクラスタリングする。よって、最初に算出される動きの精度が大局的な動きの精度に大きく寄与する。しかし、最初に求められる局所的動きにはズームなどの微小なデータが反映されない可能性²があり、検出カメラパラメータの検出精度が期待できないという問題があった。

本章では、7種類のカメラモーションと平行移動動きベクトルの関係を定式化[46][47]し、全てのカメラモーションに対応できるスプライトモデルを提案する。また、ボトムアップ的アプローチを基本として、平行移動モデルによる動きベクトルの特徴空間での分布の特徴を利用したカメラパラメータ算出法を提案する。これにより、従来技術の問題点である、アウトライヤの影響を少なくし、局所的動きベクトルの曖昧性を排除し、微妙なズーム、ロテートの検出を可能とする。さらに、カメラパラメータ算出法を用いたスプライト自動生成法を提案し、提案アルゴリズムをビデオクリップに適用して、カメラパラメータ算出実験とスプライト生成実験を行ない、スプライト生成精度の評価を行なう。

以下、3.2節では、カメラパラメータと平行移動動きベクトルの関係の定式化について検討し、スプライトモデルを提案する。3.3節では、カメラパラメータ算出アルゴリズムの提案を行なう。

²例えば CIF(352 × 288) サイズの画像において、1.003 倍という微小なズーム値はフレームの両端を比較することで検出できるが、局所的には検出不可能であることから推測できる。

表 3.1: カメラパラメータの分類

Camera location	Image scaling	Image rotation	Vertical image translation	Horizontal image translation
Still	zoom	rotate	pan	tilt
Move	doly	—	track	boom

3.4 節では、スプライト生成アルゴリズムについて述べる。3.5 節では、提案アルゴリズムを用いたカメラパラメータ算出実験を行ない、主観的に得られるカメラ操作と比較検討と MSE を用いた精度の評価を行なう。3.6 節では、スプライト生成実験を行なう。3.7 節では本章のまとめと課題について述べる。

3.2 カメラパラメータと平行移動動きベクトルの関係

3.2.1 カメラモデル

本節ではカメラパラメータと平行移動動きベクトルの関係の定式化を行う。カメラパラメータをカメラモーションを反映した画像の大局的な動きを表す変換行列の各成分と定義する。図 3.1-a にカメラモデルを示す。視点 O (View point) から見た物体 (Object) $P(X, Y, Z)$ は投影面 (Project plane) 上の点 $p(x, y)$ に投影される。

カメラの z 方向への位置の移動 Δr がカメラと点 $p(x, y)$ の実際の距離 z に比べて十分小さいという仮定の下、各カメラパラメータと平行移動動きベクトル (2 パラメータ) の関係を述べる。

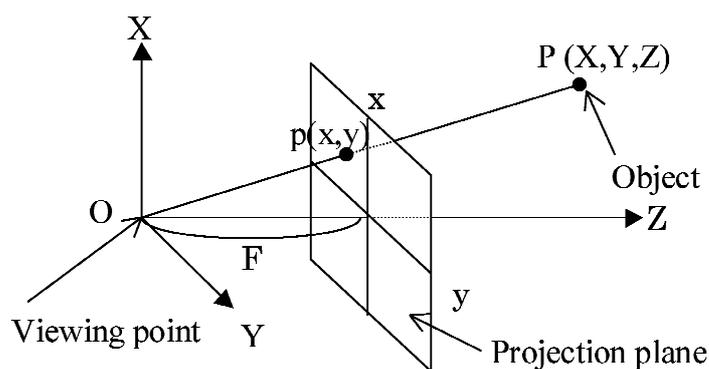
ズーム、ロテート、パン、チルトと平行移動動きベクトルの関係

撮像位置の変化しないカメラ操作 (ズーム, ロテート, パン, チルト) について, これらのカメラ操作と平行移動動きベクトルの関係は, 従来から定式化がされており [13][35][39], さまざまなグローバル動き補償法やカメラパラメータ抽出モデルで用いられている。本論文でも, 従来の定義をそのまま用いる。 A をカメラのズーム倍率, θ をカメラの回転角度とすると, 座標系 1 の点 $p1(x, y)$ から座標系 2 の点 $p2(x', y')$ への変換式は, 式 (3.1) のように表される。

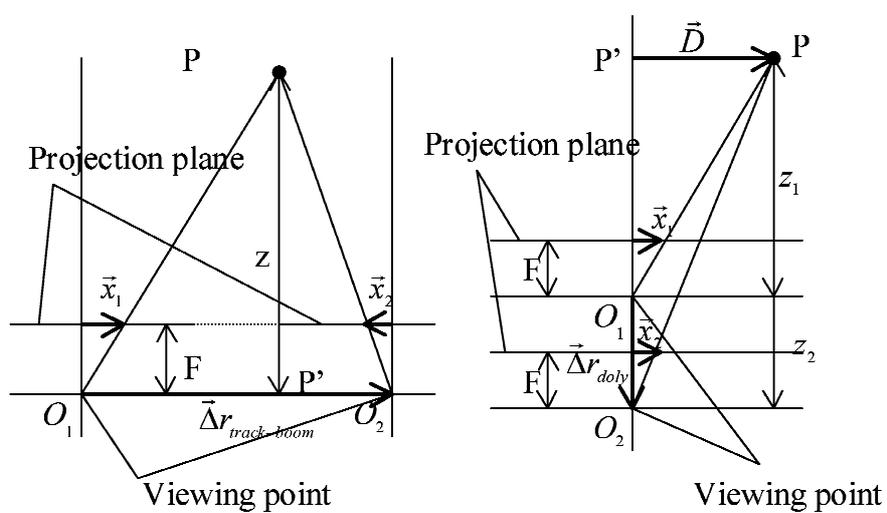
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{zoom} & b_{rotate} \\ -b_{rotate} & a_{zoom} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_{pan} \\ d_{tilt} \end{bmatrix} \quad (3.1)$$

where,

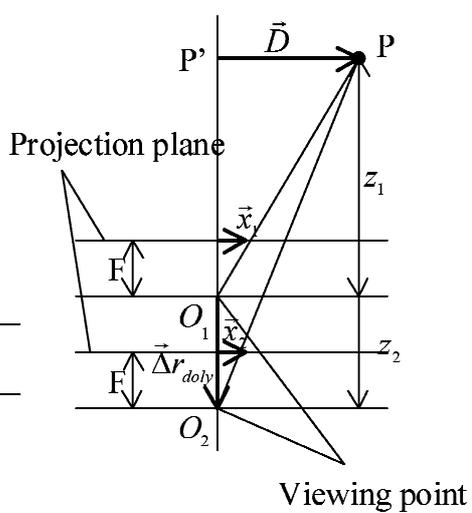
$$a_{zoom} = A \cos \theta$$



(a) Camera model.



(b) Tracking and booming model.



(c) Dolying model.

図 3.1: カメラモデル .

$$b_{rotate} = A \sin \theta$$

a_{zoom} , b_{rotate} , c_{pan} , d_{tilt} はそれぞれカメラモーションのズーム, ロテート, パン, チルトを表すスカラー値の係数 (カメラパラメータ) である.

ここで, (u, v) を x および y 方向の平行移動モデルによる動きベクトルとすると, (u, v) は座標値 x, x', y, y' を用いて, 以下の式 (3.2) のように書ける.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x - x' \\ y - y' \end{bmatrix} \quad (3.2)$$

式 (3.2) を式 (3.1) に代入すると, 以下の式 (3.3) が得られる.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a'_{zoom} & b_{rotate} \\ -b_{rotate} & a'_{zoom} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_{pan} \\ d_{tilt} \end{bmatrix} \equiv \vec{V}_1 \quad (3.3)$$

where,

$$a'_{zoom} = a_{zoom} - 1$$

すなわち, 位置座標同士の変換の場合は a_{zoom} を, 動きベクトルと座標値と間の変換の場合は a'_{zoom} をズームの係数として用いればよい. ここで, 式 (3.3) で表される, カメラの撮像位置の移動を伴わない時の動きベクトルを \vec{V}_1 とおく.

トラック, ブーム, ドリーと平行移動動きベクトルの関係の導出

撮像位置の変化するカメラ操作 (トラック, ブーム, ドリー) に関しては, 従来より透視投影モデル等により, 定式化されている. しかし, 平行移動モデルによる定式化はなされていないので, 以下にその様子を述べる. カメラ操作と図 3.1-b にトラックとブームと平行移動動きベクトルの関係を示す. 視点 (カメラの位置) が O_1 (座標系 1) から O_2 (座標系 2) に距離 $\Delta r_{track-boom}$ だけ変化したものとする. 物体の位置 P から直線 O_1O_2 に向かって垂線を下ろし交わった点を P' とする. 直線 AP' の大きさを物体の視点からの奥行き z と定義する. カメラの焦点距離を F とする. 座標系 1 での物体の投影座標を \vec{x}_1 , 座標系 2 での物体の投影座標を \vec{x}_2 とすると, 以下の式 (3.4), (3.5) が成り立つ.

$$FO_1\vec{P}' = z\vec{x}_1 \quad (3.4)$$

$$FO_2\vec{P}' = z\vec{x}_2 \quad (3.5)$$

上記式 (3.4) から式 (3.5) を減算し, $\Delta r_{track-boom} = O_1\vec{P}' - O_2\vec{P}'$ を代入すると, 算出される平行移動動きベクトル $\vec{\Delta x} \equiv \vec{x}_2 - \vec{x}_1$ は, 以下の式 (3.6) で表される.

$$\vec{\Delta x} = \frac{F}{z} \Delta r_{track-boom} \quad (3.6)$$

これは、動きベクトルの方向がカメラの移動の方向で、その大きさは物体とカメラの間の距離 z に反比例し、かつ、カメラの移動距離に正比例することを示す。

次に、図 3.1-c にドリーと平行移動動きベクトルの関係を示す。視点が O_1 (座標系 1) から O_2 (座標系 2) に距離 Δr_{doly} だけ変化したものとする。物体 P から撮像面に対して平行にカメラの光軸に対して垂線を下ろし、カメラの光軸と交わった点を P' とする。直線 $P\vec{P}' \equiv \vec{D}$ とする。直線 $O_1\vec{P}'$ の大きさを z_1 、直線 $O_2\vec{P}'$ の大きさを z_2 とする。カメラの焦点距離を F とする。座標系 1 での物体の投影座標を \vec{x}_1 、座標系 2 での物体の投影座標を \vec{x}_2 とすると、式 (3.7)、(3.8) が成り立つ。

$$\vec{x}_1 = \frac{F}{z_1} \vec{D} \quad (3.7)$$

$$\vec{x}_2 = \frac{F}{z_2} \vec{D} \quad (3.8)$$

ここで、 $\Delta r_{doly} \equiv z_2 - z_1$ とすると、動きベクトル $\vec{\Delta x} \equiv \vec{x}_2 - \vec{x}_1$ は、上記式 (3.7)、(3.8) から式 (3.9) のように書ける。

$$\begin{aligned} \vec{\Delta x} &= \vec{x}_2 - \vec{x}_1 \\ &= F \frac{z_2 - z_1}{z_1 z_2} \vec{D} \\ &= F \frac{\Delta r_{doly}}{z_1 z_2} \vec{D} \end{aligned} \quad (3.9)$$

さらに、式 (3.9) は $z_1 \approx z_2 \equiv z$ とおいて、座標値 \vec{x}_1 を用いると式 (3.10) のように書ける。

$$\vec{\Delta x} = \frac{\Delta r_{doly}}{z} \vec{x}_1 \quad (3.10)$$

これは、動きベクトルの方向が座標値の方向で、その大きさは物体とカメラの間の距離 z に反比例し、かつ、カメラの移動距離に正比例することを示す。

よって、式 (3.6)、(3.10) より、トラック、ブーム、ドリーを反映した場合の平行移動動きベクトル (u, v) は式 (3.11) のように書ける。

$$\begin{aligned} \begin{bmatrix} u \\ v \end{bmatrix} &= \frac{1}{z} \begin{bmatrix} a'_{doly} & 0 \\ 0 & a'_{doly} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{z} \begin{bmatrix} c_{track} \\ d_{boom} \end{bmatrix} \\ &\equiv \vec{V}_2 \end{aligned} \quad (3.11)$$

where,

$$a'_{doly} = \Delta r_{doly}$$

$$a'_{doly} = a_{doly} - 1$$

$$\begin{bmatrix} c_{track} \\ d_{boom} \end{bmatrix} = F \Delta r_{track-boom}$$

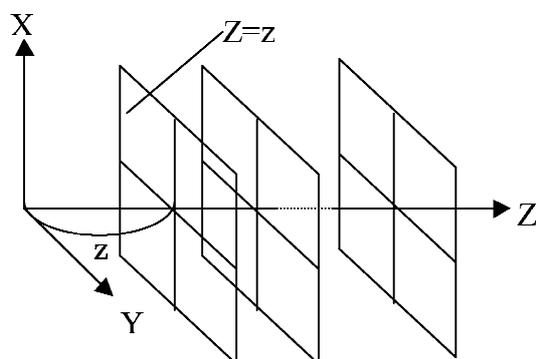


図 3.2: スプライトモデル.

ここで, a_{doly} , a'_{doly} , c_{track} , および d_{boom} はそれぞれカメラモーションのドリー, トラック, ブームを表すスカラー値の係数 (カメラパラメータ) である. ズームと同様, 座標系の変換には a_{doly} (a'_{doly} に 1 を加算した値) を, 動きベクトルと座標の間の変換には a'_{doly} を用いる. z は点 (x, y) における実際の奥行き値である. ここで, 式 (3.11) で表されるカメラの撮像位置の変化のない場合の動きベクトルを \vec{V}_2 とおく.

全カメラパラメータと平行移動動きベクトルの関係

点 (x, y) における動きベクトルは式 (3.3), (3.11) で示すカメラモーションによる動きベクトル \vec{V}_1 , \vec{V}_2 の線形結合で表される. その様子を式 (3.12) に示す.

$$\begin{aligned}
 \begin{bmatrix} u \\ v \end{bmatrix} &= \vec{V}_1 + \vec{V}_2 \\
 &= \begin{bmatrix} a'_{zoom} & b_{rotate} \\ -b_{rotate} & a'_{zoom} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_{pan} \\ d_{tilt} \end{bmatrix} \\
 &+ \frac{1}{z} \left\{ \begin{bmatrix} a'_{doly} & 0 \\ 0 & a'_{doly} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_{track} \\ d_{boom} \end{bmatrix} \right\}
 \end{aligned}
 \tag{3.12}$$

右辺第 1 項, 2 項はカメラの位置の移動がないものである. 右辺第 3 項, 4 項はカメラの位置の移動を伴う動きベクトルを反映している. また, 画像の変形に着目すると, 第 1, 3 項は画像の拡大縮小と回転の効果を表すもの, 第 2, 4 項は画像の移動の効果を表すものである. そこで, ズーム, ドリーを拡大縮小パラメータ, ロテートを回転パラメータ, パン, チルト, トラック, ブームを移動パラメータと呼ぶことにする.

3.2.2 スプライトモデル

一般にスプライトは領域の動きが平行移動モデル (2 パラメータ), アフィンモデル (6 パラメータ), 透視投影モデル (8 パラメータ) のいずれか, 即ち平面モデルで表せられる時に生成可能である [1]. 前節の式 (3.12) は $(x, y)^T$ の項で括ると式 (3.13) の様に書ける.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a(z) & b(z) \\ -b(z) & a(z) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c(z) \\ d(z) \end{bmatrix}$$

where,

(3.13)

$$\begin{aligned} a(z) &= a'_{zoom} + \frac{1}{z} a'_{doly} \\ b(z) &= b_{rotate} \\ c(z) &= c_{pan} + \frac{1}{z} d_{track} \\ d(z) &= c_{tilt} + \frac{1}{z} d_{boom} \end{aligned}$$

式 (3.2) において, 同じ奥行きにおいては 4 パラメータのアフィンモデル (ヘルマートモデル) で表されることが分かる. よって図 3.2 に示すように, 同じ奥行き毎に 1 枚のスプライトを形成できる. ビデオクリップに属する各画素がどの奥行きに属するかが分かれば, 現実世界の配置に基づいた各フレームの領域の対応が分からなくても, スプライトを生成できる. 従来のボトムアップ的アプローチ [9][10] では, 撮像位置が変化したものに対しては, 実世界の 3 次元構成に基づく各フレームの領域の対応が既知でない場合はスプライトを定義できなかった. 一方, 本モデルでは再サンプリングされた相対的奥行き毎に「ついたて」状にスプライトを配置し, 各フレームの各画素を「ついたて」の上にマップすることで上記の条件の下でもスプライトを生成できる. ビデオクリップの各画素がどのスプライトにマップされるか, すなわち, どの奥行きに相当するのかの判別は, 次節のカメラパラメータ算出アルゴリズムの中で述べる.

3.3 カメラパラメータ算出アルゴリズム

本節では前述のカメラモデルによる隣接フレーム間のマクロブロック毎に算出される平行移動動きベクトルからカメラパラメータを算出するアルゴリズムを述べる. 動きベクトルはマクロブロック単位に求められたものを用いる. 本アルゴリズムの特長は以下の点にある.

- 7 種類全てのカメラパラメータを算出できる.
- 平行移動動きベクトルを利用するので MPEG 等の既存の符号化データを利用可能である.
- 動きベクトルの偏微分値の特徴空間での分布特性を利用することで, カメラモーション以外の個々の物体の動きに影響されることなくカメラパラメータを抽出可能である.

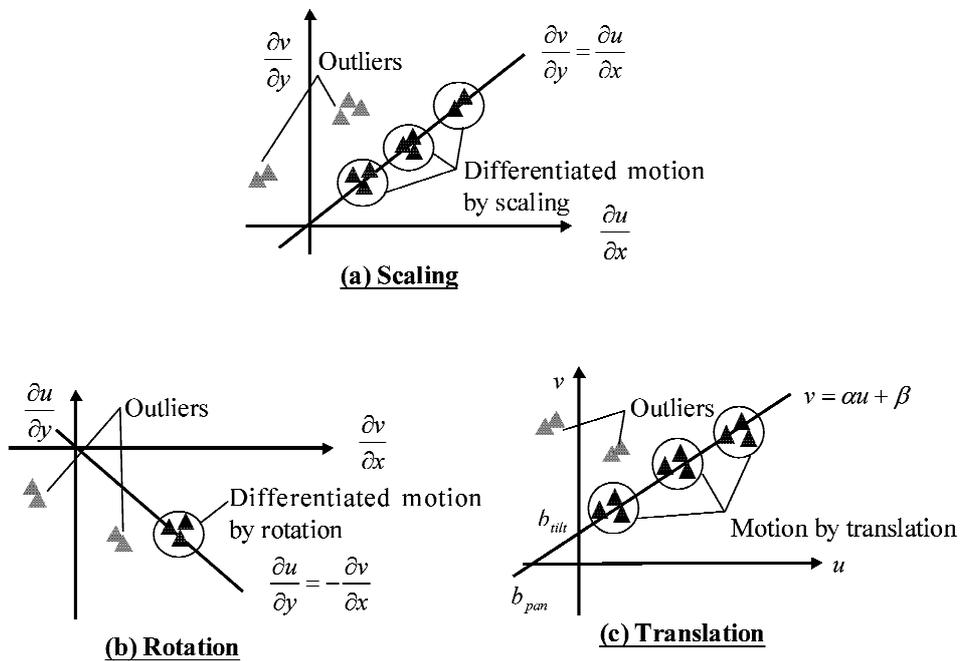


図 3.3: 平行移動動きベクトルの特徴空間での分布 .

3.3.1 拡大縮小パラメータの算出

前出の式 (3.12) の u および v を位置座標 x 及び y で偏微分すると以下の式 (3.14) が得られる.

$$a'_{zoom} + \frac{a'_{dofy}}{z} = \frac{\delta u}{\delta x} = \frac{\delta v}{\delta y} \tag{3.14}$$

式 (3.14) における偏微分値はカメラの拡大縮小パラメータ (ズーム, ドリー) と個々の動きの偏微分値を反映している. 任意の隣接フレーム間において算出されたマクロブロック毎の動きベクトルを式 (3.14) にのっとって偏微分し, これらの偏微分値を特徴空間にマップする. この様子を図 3.3-a に示す. 拡大縮小のカメラパラメータのみ反映している偏微分値は必ず同じ象限に現れ, かつ, 直線 $\frac{\delta u}{\delta x} = \frac{\delta v}{\delta y}$ 上に分布する. 一方で, 個々の動きのあるものはアウトライヤとしてそれ以外の部分に分布する. とここで, 実際の偏微分値は, 隣接間動きベクトル同士はほとんど差が出ず, 特徴空間上の座標 (0, 0) にピークが立つことが予想される. そこで, 近傍ブロック間で差分をとることを避けて N ブロック離れた動きベクトルの組み合わせにおいて差分を算出し, ブロック数 N で除して単位ブロック当たりの差分を求め, それを特徴空間にマップする. 次に, 特徴空間にマップされた動きベクトルの偏微分値をクラスタリングする. 特徴空間上の直線 $\frac{\delta u}{\delta x} = \frac{\delta v}{\delta y}$ から遠くはなれたところに分布する偏微分値はアウトライヤとして無視して, 直線付近に分布するデータのみ k 平均法などのクラスタリングを施してセントロイドを求める.

ここで、式 (3.14) から同じ距離 z に所属する微分値は同じセントロイドのクラスタに所属する。言い換えれば、相対的距離毎の「ついたて上に」実世界が再サンプリングされる。クラスタが1つの場合は相対距離が1つしか存在しないことになる。すなわち、カメラの撮像位置の移動がない場合（ズームのみ）である。特殊な場合として、このセントロイドが特徴空間の原点 $(0,0)$ にある場合は全く拡大縮小の操作がない場合である。一方、クラスタが n 個（複数）の場合、それぞれのセントロイドが相対距離を代表する。よって、カメラの撮像位置の移動があると考えられる。この場合は、ドリーのみカメラ操作、もしくはドリーとズームが同時に操作されている場合が考えられる。ドリーとズームの厳密な分離は困難である。しかし、スプライト生成においては、相対距離毎に拡大縮小パラメータ、回転パラメータ、 x および y 方向への移動パラメータの4種類が算出されれば、スプライトを構築することが可能であるので、ドリーとズーム、トラックとパン、ズームとチルトの分離がされなくてもよい。

上記の原理により、特徴空間でのクラスタの個数ならびに分布特性を調べることで、以下のように拡大縮小パラメータを算出する。

1. 強いピークが座標 $(0,0)$ 付近に現れる時は拡大縮小パラメータは無いとする。即ち、 $a'_{zoom} = a_{doly} = 0$ である。
2. 単峰性のクラスタが座標 $(0,0)$ 以外に現れる時はそのセントロイドを唯一のズーム値 a_{zoom} とし、ドリーの値はない、即ち、 $a_{doly} = 0$ とする。
3. 多峰性のクラスタ分布が現れる場合はドリーを含むカメラモーションとする³。

3.3.2 回転パラメータの算出

同様にして、前出の式 (3.12) の u および v を位置座標 y 及び x で偏微分すると以下の式 (3.15) が得られる。

$$b_{rotate} = \frac{\delta u}{\delta y} = -\frac{\delta v}{\delta x} \quad (3.15)$$

式 (3.15) における微分値はロテートと個々の動きの偏微分値を反映した値を示す。ロテートのみ反映する場合は、図 3.3-b に示すように、同じ象限の直線 $\frac{\delta u}{\delta y} = -\frac{\delta v}{\delta x}$ 上に分布し、個々の動きを含む場合はそれ以外の場所に分布する。これも前小節同様、実際に算出する場合は N ブロック離れた動きベクトルのすべての組合せにおいて差分を求め、 N で除した値を特徴空間にマップしてアウトライヤは無視してクラスタリング処理を行なう。この場合は、必ず単峰性の強いピークが現れるので、これの唯一のセントロイド（重心）をロテートの値とする。

³スプライト生成においては、ドリーとズームの混在する拡大縮小パラメータとして a_{mix} を定義し、この値を使用する。

3.3.3 移動パラメータの算出

式 (3.12) を変形すると式 (3.16) が得られる．各々のブロック毎に，前小節で算出された拡大縮小パラメータ V_{scale} と回転パラメータ V_{rotate} を式 (3.16) に代入して，元の動きベクトル V_{all} から拡大縮小パラメータと回転パラメータを差し引いた平行移動動きベクトル V_{trans} を算出する．

$$\begin{aligned} & \begin{bmatrix} b_{pan} \\ b_{tilt} \end{bmatrix} + \frac{1}{z} \begin{bmatrix} b_{track} \\ b_{boom} \end{bmatrix} \\ & = V_{all} - V_{scale} - V_{rotate} \equiv V_{trans} \end{aligned} \quad (3.16)$$

これらの動きベクトルを k 平均法などによってクラスタリングすると図 3.3-c の様にクラスタの分布が現れる．カメラモーションを反映する動きベクトルは任意の直線上に乗り，かつ，同じ象限にある．

拡大縮小パラメータの算出時と同様，式 (3.16) から同じ距離 z に所属する微分値は同じセントロイドのクラスタに所属する．クラスタの数分の相対距離に実世界が再サンプリングされたことになる．クラスタが 1 つの場合はカメラの撮像位置の移動がない場合（パン，チルトのみ）と考えられる．さらに，クラスタが n 個（複数）の場合，それぞれのセントロイドが相対距離を代表するので，カメラの撮像位置の移動があると考えられる．すなわち各セントロイドはパン，チルト，トラック，ブームの加算値を反映している．カメラ操作によるクラスタは特徴空間上のセントロイドと原点の距離がこの相対距離に比例するように配置されるので，任意の直線 $v = \alpha u + \beta$ 上に形成される．この直線から遠く離れているクラスタはアウトライヤと見なすことができる．相対距離によって値が変わるのはトラック，ブームの値なので，全クラスタに共通のパン，チルトの値はそれぞれ直線 $v = \alpha u + \beta$ の x 切片， y 切片で与えられる．

上記の原理により，クラスタの分布の様子から移動パラメータは以下のようにして求められる．

1. 単峰性ならば，移動パラメータはパン・チルトのみとなり，値は式 (3.17) になる．

$$\begin{bmatrix} b_{pan} \\ b_{tilt} \end{bmatrix} = \begin{bmatrix} C_x \\ C_y \end{bmatrix} \quad (3.17)$$

ここで， $(C_x, C_y)^T$ は特徴空間におけるセントロイドである．

2. 多峰性のクラスタが形成される場合，式 (3.18) で表されるように直線の切片がパン，チルトの値であり， n 個のセントロイドの値 $(C_x^n, C_y^n)^T$ からパン・チルトの値を引いたものがトラック・ブームの値となる．

$$\begin{bmatrix} b_{track}^n \\ b_{boom}^n \end{bmatrix} = \begin{bmatrix} C_x^n - b_{pan} \\ C_y^n - b_{tilt} \end{bmatrix} \quad (3.18)$$

ここで，ドリー，トラック，ブームの算出時に見られる n 個のセントロイドはカメラからの相対距離を反映している．即ち，実空間をカメラからの距離毎に n 個にサンプリングする．

3.4 スプライト生成アルゴリズム

前節よりカメラの位置が固定の場合は唯一のスプライトが生成され、カメラの位置が変動した場合はサンプリングされた奥行き毎にスプライトが生成される。当節では、この2つの場合についてスプライトを自動生成するアルゴリズムを提案する。

3.4.1 カメラの位置が固定の場合

前節より、カメラの位置が固定の場合、考えられるカメラパラメータはパン、チルト、ズーム、ロテートの4種類である。これらのカメラパラメータは一組のヘルマート変換の成分を表すものであり、一枚の平面を記述できる。以下にこの場合のスプライト生成アルゴリズムを説明する。

隣接フレーム間カメラパラメータ算出

隣接フレーム間で平行移動モデルによる動きベクトルを算出する。その動きベクトルを前節のカメラパラメータ算出アルゴリズムに適用して隣接フレーム間のカメラパラメータを算出する。

基準フレームからのカメラパラメータ算出

ここで、座標系 n を $(x_n, y_n)^T$ 、座標系 $n-1$ を $(x_{n-1}, y_{n-1})^T$ とすると座標系 n から座標系 $n-1$ への変換は式 (3.19) の様に見える。

$$X_n = GX_{n-1} + H \quad (3.19)$$

where,

$$X_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$$

$$G = \begin{bmatrix} a_{zoom} & b_{rotate} \\ -b_{rotate} & a_{zoom} \end{bmatrix}$$

$$H = \begin{bmatrix} c_{pan} \\ d_{tilt} \end{bmatrix}$$

ここでズーム値は a_{zoom} を用いる。

ヘルマート変換は、座標系 1 と座標系 2、かつ、座標系 2 と座標系 3 の間がヘルマート変換で表される時、座標系 1 と座標系 3 の間もヘルマート変換で表されるという性質を持つ。即ち、各座標間の変換行列の各成分は常に同じカメラパラメータを表す。よって、時間的に各行列成分を追うことで、カメラの動きを把握可能である。

隣接フレーム間カメラパラメータが全く理想的に算出された場合、任意の座標系 X_n 上の点は式 (3.20) で表される漸化式の様に基準座標 X_0 上へマップされる。

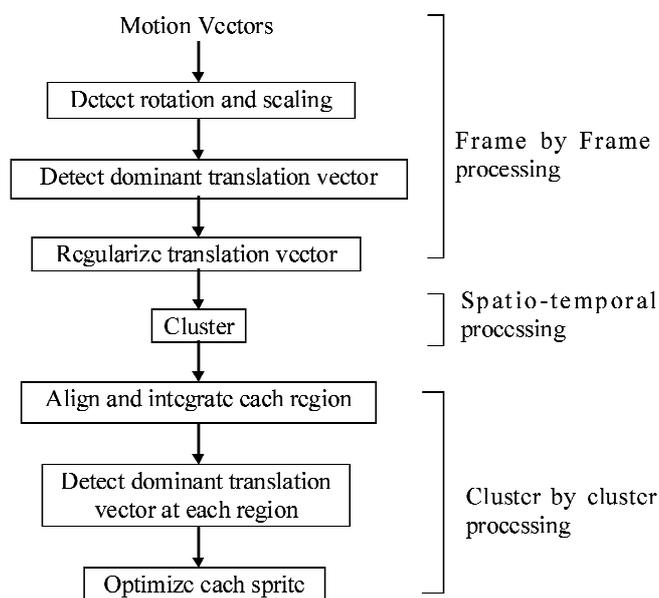


図 3.4: スプライト生成アルゴリズム .

$$X_0 = G_{n-1}^{tmp} G_n X_n + G_{n-1}^{tmp} H_n + H_{n-1}^{tmp} \quad (3.20)$$

where,

$$G_1^{tmp} = G_1$$

$$H_1^{tmp} = H_1$$

$$G_2^{tmp} = G_1 G_2$$

$$H_2^{tmp} = G_1 H_2 + H_1$$

$$G_n^{tmp} = G_{n-1}^{tmp} G_n$$

$$H_n^{tmp} = G_{n-1}^{tmp} H_n + H_{n-1}^{tmp}$$

こうして求められた基準フレームからの変換を絶対カメラパラメータと呼ぶことにする .

フレームの統合

前小節で基準フレームにマップされた各々のフレームは、同じ座標に属する画素群のメディアン値や平均値で代表する [9][10]、あるいは上書きを繰り返すことで時間方向に統合し、枚のスプライトを生成する .

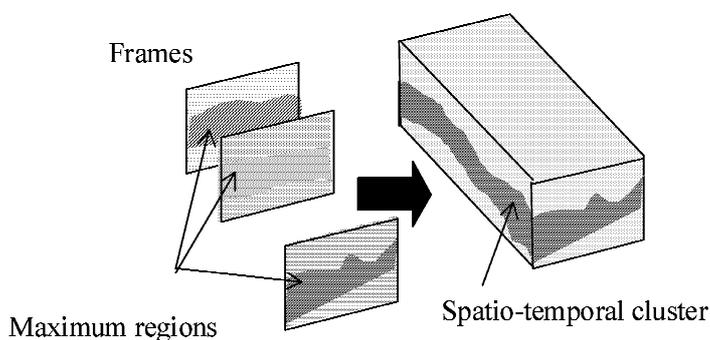


図 3.5: ビデオクリップ内処理 .

3.4.2 カメラの位置の移動がある場合

カメラの位置が移動する場合はサンプリングされた奥行き（相対的な奥行き）毎にスプライトが生成されるためにビデオクリップに属する画素群を相対的な奥行き毎にクラスタリングする必要がある．ここでは，条件として，カメラの移動がほぼ定速な場合を仮定する．手順を図 3.4 に示し，以下に詳細を述べる．

フレーム毎処理

前節の式 (3.14)，(3.15) にのっとり，動きベクトルの偏微分値を用いて拡大縮小パラメータと回転パラメータを算出し，式 (3.16) によって移動のみを反映した動きベクトル（以下，移動ベクトル）をマクロブロック毎に算出する．ここで算出された移動ベクトルは奥行きの逆数を反映するものである．

ビデオクリップ内処理

図 3.5 にビデオクリップ内処理のイメージを示す．ここでカメラの動きがほぼ定速の場合は，各々のフレーム内で最大面積を占める大きさを持つ移動パラメータは実空間ではカメラからの距離はほぼ等しいという仮説を立てて各フレーム毎の領域の対応をとる．即ち，各フレームの該当領域は同じ奥行きのスプライトにマッピングされる．各フレームの全てのブロックの移動パラメータの大きさを算出し，最大面積をもつ移動パラメータの大きさを基準値として，各フレーム内の移動パラメータの大きさを正規化する．さらに，正規化された大きさを特徴とする 1 次元の特徴空間内において，ビデオクリップ内の全ての移動ベクトルの大きさの正規化値をクラスタリングする．各フレームにおいて， n 個のセントロイドに所属する n 個の移動パラメータを式 (3.18) によって算出する．

次にビデオクリップ内の画素がどの奥行きのスプライトに所属するかを決定する． n 個の奥行

表 3.2: 使用した画像 (1)

Image	Type	Frames	Basical frame number.
MIT sequence	CIF	1-150	1
Stefan	CIF	1-300	1
Flower garden	CIF	50-110	1

表 3.3: 使用した画像 (2) .

Image	Type	Frames	Basical frame No.
MIT sequence	CIF	1-75	1
Stefan	SIF	200-300	200
Flower garden	CIF	50-110	50

きを反映する n 組の移動パラメータと前出の回転パラメータ, 拡大縮小パラメータを用いて, 前フレームから n 個の予測画像を作成する. 着目画像と予測画像の対応する $m \times m$ 画素の小ブロックにおいて, 2乗誤差が最小となるカメラパラメータの組をその小ブロック内の画素の所属する奥行きと決定する. 全てのフレームにおいて同様の処理を行ない, 全ての画素の所属する奥行きを決定する.

奥行き毎処理

上記で算出したフレーム毎の n 組の隣接フレーム間カメラパラメータを用いて, 式 (3.20) によって奥行き毎の絶対カメラパラメータを算出する. カメラの移動のない場合と同様の絶対カメラパラメータ算出と時間方向への統合処理を各奥行き毎に行なう.

3.5 カメラパラメータ算出実験

前節のカメラパラメータ算出アルゴリズムによってカメラパラメータ算出実験を行なった.

3.5.1 実験

表 3.2 に示す標準画像を用いて以下の手順にてカメラパラメータ算出実験を行なった.

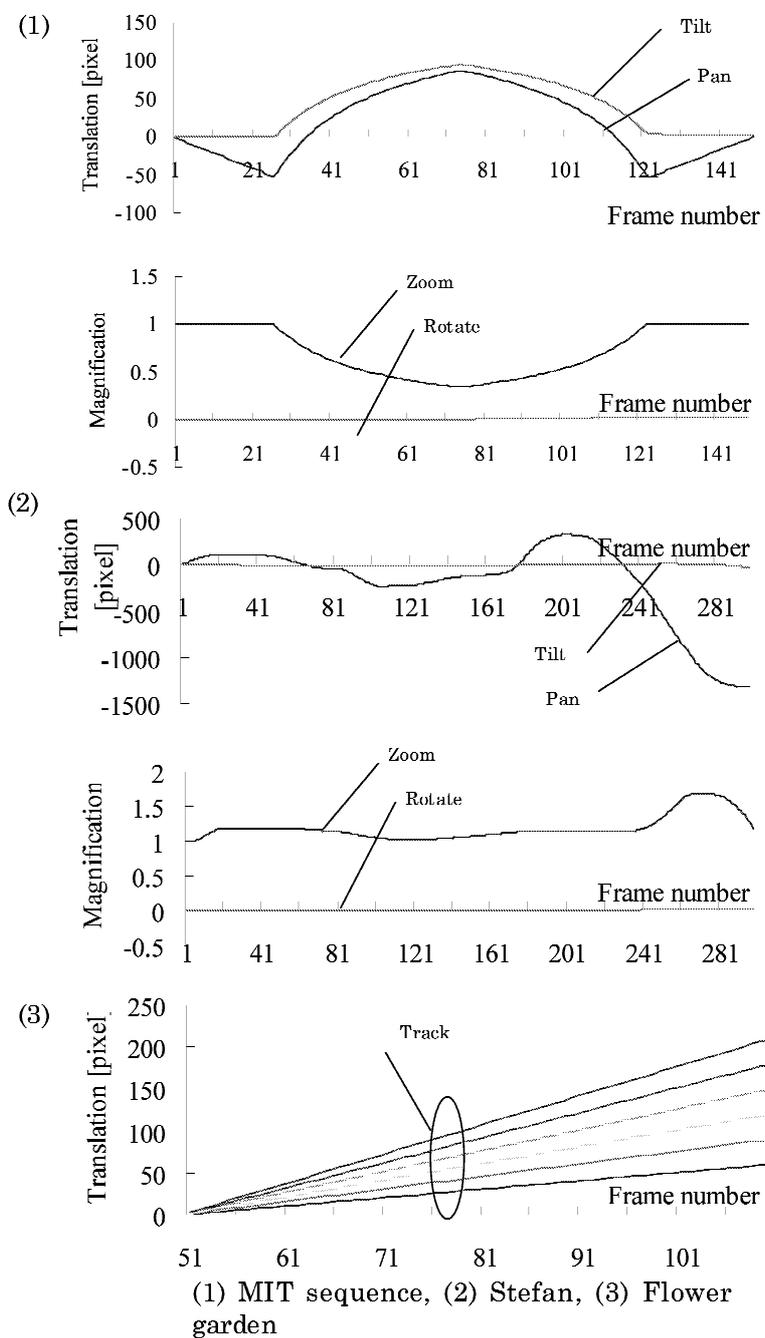
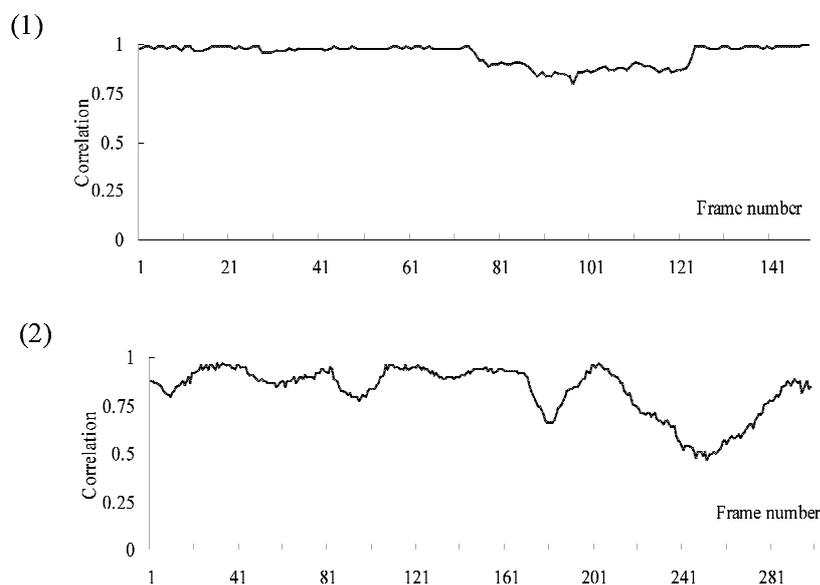


図 3.6: 絶対カメラパラメータ.



(1) MIT sequence, (2) Stefan

図 3.7: 原画像と予測画像の相関 .

1. 各々の画像において、フレーム間のカメラパラメータを算出した .
2. カメラパラメータの算出精度を測定するために、“MIT sequence” , “Stefan” に対して、求められたフレーム間カメラパラメータを用いて予測画像を作り、次のフレームの画像と相関を用いて比較を行なった .⁴
3. 各々の画像において、最初のフレームを基準フレームとしてフレーム間カメラパラメータを式 (3.19) , (3.20) に代入して得られる絶対カメラパラメータを算出し、その波形を画像から目視により認識できるカメラモーションと比較した .

3.5.2 結果並びに考察

算出された絶対カメラパラメータを図 3.6 に示す . 但し、ズームには座標変換係数である a_{zoom} の値を用いる . ここで、表 3.4 に原画像から目視によって得られる実際のカメラ操作をまとめた . また、“MIT sequence” , “Stefan” における、予測画像との相関を図 3.7 に示す .

図 3.6 の絶対カメラパラメータ値は基準座標からカメラがどのように振れているかを示すもの

⁴オブジェクト単位にカメラパラメータの算出精度を測定することを目的とするため、単一背景スプライトが生成される“MIT sequence” と“Stefan” で十分とし、多数の領域に分割される“Flowergarden” に関しては各領域のカメラパラメータの精度が上記の 2 つの結果から類推できるとし、ここでは割愛した .

表 3.4: 各ビデオクリップにおけるカメラ操作 .

Image	Frames	Camera operations
MIT sequence	1-25	pan-left
	25-75	pan-right, tilt-down, zoom-in
	75-125	pan-left, tilt-up, zoom-out
	125-150	pan-right
Stefan	1-20	pan-right, zoom-out
	20-104	pan-left
	104-186	pan-right
	186-300	pan-left, zoom-out, zoom-in
Flower garden	50-110	track-right

である．移動 (Translation) を表すグラフにおいて，傾きが正の時は右方向 (right)，負の時は左方向 (left) のカメラ操作がある．また，拡大縮小 (Scaling) および回転 (Rotation) を表すグラフにおいて，傾きが正の場合は縮小および時計回りの回転を表し，傾きが負の場合は拡大および反時計回りの回転を表す．”MIT sequence” はもともと一枚の静止画像からカメラパラメータによって切り出してきた画像で，シーケンスの中央 75 フレーム付近を境に全く逆のカメラ操作を行なっている．”Stefan” はテニス選手が左右にボールを追いかける様子をカメラで追った画像である．”Flower garden” は景色をカメラの位置を移動させながら撮影したものである．”Flower garden” においては奥行き方向に 6 つの領域にサンプリングされ，各々の領域がその距離に反比例して移動していく様子が分かる．これらの目視から得られるカメラ操作 (表 3.4) と図 3.6 のグラフの波形から，提案アルゴリズムによって目視に近いカメラ操作が抽出できていることが分かる．

また，図 3.7 より，カメラモーションがズームアウトのところ，もしくは個々の動きが大きいところは相関値が低い．ズームアウト時では，新たな領域の出現があるので，カメラの予測精度が落ちると考えられる．それ以外の部分は相関 0.9 以上であり，隣接間カメラパラメータが良好に抽出できていることが分かる．

3.6 スプライト生成実験

前節のスプライト生成アルゴリズムを用いて以下のスプライト生成実験を行なった．使用した画像を表 3.3 に示す．

MIT sequence 1-75 frames



図 3.8: 生成されたスプライト”MIT sequence” .

Stefan 200-300 frames

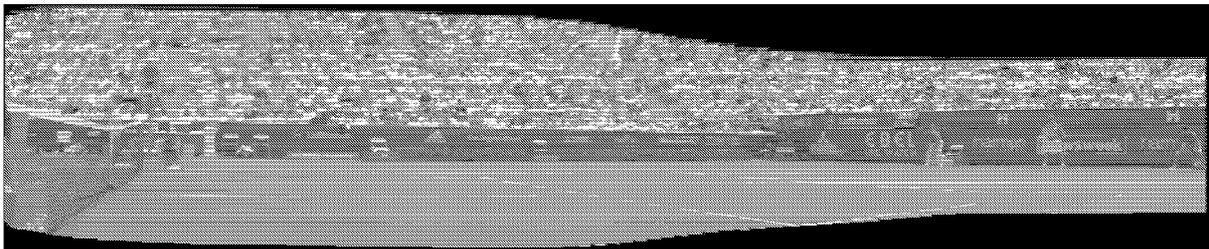


図 3.9: 生成されたスプライト”Stefan” .

3.6.1 実験

以下の手順にて実験を行なった .

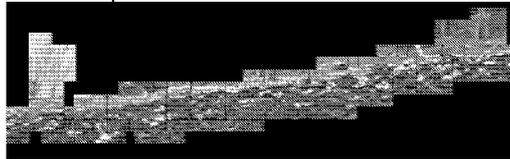
1. 各々の画像に関してスプライトを生成した .
2. “MIT sequence” に関してはフレーム毎にスプライトから予測画像を切り出して相関値を用いて原画像との比較を行なった .

Flower garden 50-110 frames

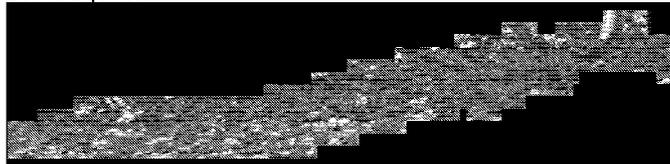
First sprite



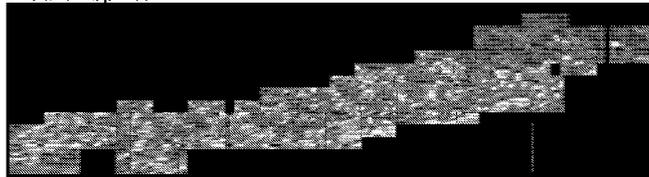
Second sprite



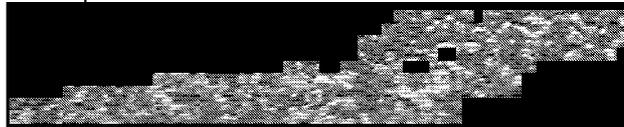
Third sprite



Fourth sprite



Fifth sprite



Sixth sprite

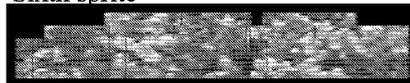
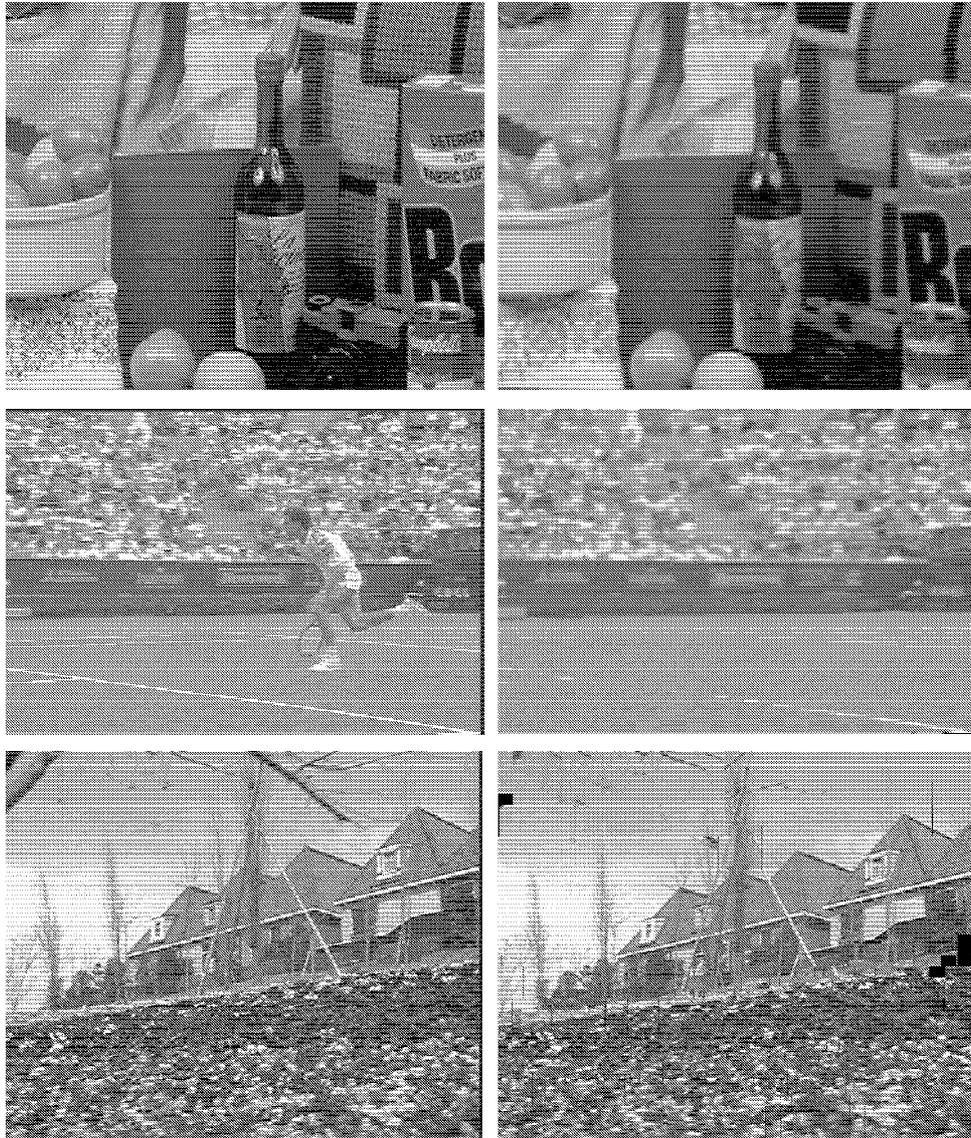


図 3.10: 生成されたスプライト”Flower garden” .



Upper: MIT sequence 50th frame

Middle: Stefan 250th frame

Lower: Flower garden 55th frame

*Left: original image, right: extracted image

図 3.11: 背景の再生画像の例 .

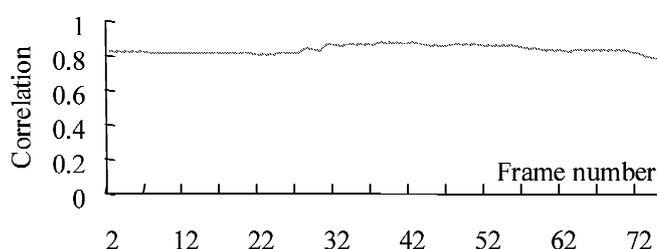


図 3.12: 背景の再生画像と原画像の相関 (MIT sequence) .

3.6.2 結果並びに考察

各々の画像に対する生成スプライトを図 3.8, 図 3.9, 図 3.10 に示す. また, 図 3.11 に各々のスプライトから絶対カメラパラメータを用いて, 再構成された画像の例を示す. また, "MIT sequence" におけるスプライトから再構成された画像と原画像の相関を図 3.12 に示す.

図 3.8, 図 3.9, 図 3.10 より, カメラモーションのみを反映する背景領域がスプライトとして視覚的に良好に抽出できていることが分かる. 特に "MIT sequence" においては, ビデオクリップ作成のもととなった 1 枚の静止画像がスプライトとして抽出されている. "Stefan" において, スプライトから絶対カメラモーションによって切り出された画像はテニス選手のいない画像となっている. これは, 時間方向への統合にメディアン値を用いているため, 時空間方向にアラインメントされた画素のうち, 半数以上が背景画素が占めているとスプライトに背景画素が選択されるからである. "Flower garden" においては, 初期クラスタ中心を 6 つに設定したので, 奥行き毎に 6 つの領域にサンプリングされ, 6 つのスプライトが生成されている.

図 3.11 において, 切り出した画像と原画像を比較してみると, 原画像に近い切り出し画像が生成されていることがわかる. さらに図 3.12 において, 理想的なパン・ズーム系である "MIT sequence" の切り出し画像と原画像の相関係数がおもむね 0.8 以上と, 良好にスプライトを生成できていることが分かる.

カメラパラメータおよびスプライトの抽出精度については, 実際の撮影時のカメラ操作の定量的なデータが不明であるので具体的な評価はできないが, 視覚的に良好な結果を得ているので, スプライトを使った超低レート符号化には十分利用できると思われる.

3.7 本章のまとめ

本章ではビデオクリップを撮像した時のカメラモーションに着目し, カメラモーションを反映した背景画像をスプライトとして生成することを目的として, (1) カメラモーションと平行移動動きベクトルの関係の定式化, (2) カメラパラメータ算出アルゴリズムの提案, (3) スプライト自動生成アルゴリズムの提案を行ない, 実際にアルゴリズムにビデオクリップを適用した結果, 直観的にカメラモーションに近いパラメータを抽出でき, カメラモーションのみを反映した背景画

像をスプライトとして良好に抽出できることが分かった。

しかし、今回の検討では隣接フレーム間カメラパラメータを輝度値を用いて検定を行っていないので、平坦な色使いの画像の場合、動きベクトルがゼロの値に集中することが予測され、隣接フレーム間カメラパラメータが誤抽出されることが考えられる。また、絶対カメラパラメータの算出においては、座標間の変換式 (3.19), (3.20) が全く誤差を含まない理想の状態を仮定しているため、基準フレームから離れたフレームでは絶対カメラパラメータは誤差が累積することが予測される。そこで今後は、隣接フレーム間カメラパラメータ算出時における何らかの検定法と、絶対カメラパラメータ算出時の誤差の吸収法の確立が望まれる。さらに、完全にアウトライヤの影響を除去したカメラパラメータを算出するには前景となるオブジェクトを切り出し、背景部分のみでカメラパラメータを推定する必要がある。そこで、オブジェクトの抽出とリンクしたカメラパラメータ推定とスプライト生成が課題である。

第4章 グローバルモーション算出法の改良と各符号化方式への適用

4.1 はじめに

オーディオビジュアル符号化の次世代標準 MPEG-4[3] では新たにグローバル動き推定 (Global Motion Estimation: GME) やスプライトを用いた符号化方式の検討がなされている。GME やスプライト符号化では、画面全体の動きを一組の動きパラメータで表すためのグローバルモーション (GM: Global Motion) 算出技術が中核をなしている。一般的な GM 算出法では、各々のマクロブロック単位の動きと GM の差分の二乗和が最少になるように GM が算出される [19]。

一方、スプライト符号化において、カメラの動きのみを反映した背景スプライトを生成する場合、アウトライヤ (カメラモーション以外の物体の動き等の外れデータ) の影響が無い GM を算出する必要がある。スプライト生成を目的とした GM の算出においては以下の 2 つの方法が考案されてきた。一つはあらかじめ算出された局所的な動きベクトルをクラスタリングし、最大クラスタのセントロイドを GM とするボトムアップ的な手法である [11][12]。もう一つは画面全体を変形して直接的に GM を算出するトップダウン的な手法である [9][10]。前者は、輝度勾配のなだらかな画像では局所的な動きが正確に算出できないという問題があった。また、最大クラスタがインライヤ (カメラモーションのみを反映したデータ) を代表しているとは限らなかった。一方、後者は、原画像と予測画像を用いて算出された差分画像の全体の画素の 2 乗和を最小とする値が GM として算出される。この方法は輝度勾配の問題は回避できるものの、動きベクトルの平均値が GM の値となるので、アウトライヤを除いた GM の算出は困難であった。

本章では、ボトムアップ的な方法をベースに、輝度勾配の問題を回避し、アウトライヤの影響の少ない GM 算出アルゴリズムを提案する。また、提案方式より得られた GM を用いて、MPEG-4 Ver.11[48] における (1)LME 符号化、(2)GME 符号化、(3)ダイナミックスプライト符号化、(4)スタティックスプライト符号化を行い、それぞれの符号量と画品質を評価する。以降、4.2 節では GM 算出アルゴリズムの改良、4.3 節では改良アルゴリズムを用いたグローバルモーション算出実験、4.4 節では画像符号化への適用、4.5 節では本章のまとめと課題について述べる。

4.2 GM 算出アルゴリズムの改良

本アルゴリズムでは、式 (4.1) に示す 4 パラメータのヘルマート変換を用いた GM モデル [47][50] を使用する。ヘルマート変換 (4 パラメータ) はアフィン変換 (6 パラメータ) の特殊な場合で

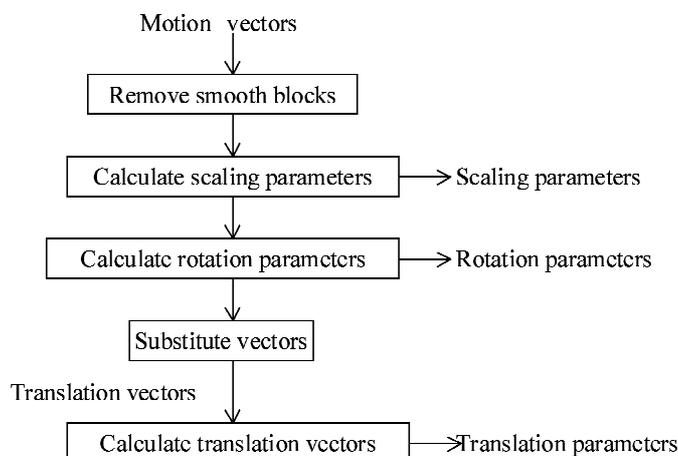


図 4.1: GM 算出アルゴリズムのながれ .

あり¹, パン, チルト, ズーム, 回転といったカメラ操作と親和性がよく, カメラで撮像された背景画像からスプライトを生成する場合に過不足無いパラメータ構成となっている .

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a' & b \\ -b & a' \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \quad (4.1)$$

$$a = a' + 1$$

4 パラメータ a, b, c, d はそれぞれ, 拡大縮小, 回転, 平行移動 (x 方向および y 方向) を表す . また, (u, v) は任意の座標 (x, y) での平行移動モデル (2 パラメータ) による動きベクトルを表している .

図 4.1 に GM 算出のアルゴリズムを示し, 以下にその詳細を示す . 本アルゴリズムは基本的にクラスタリングの結果のセントロイドを GM とするボトムアップ的アプローチをとる . 動きベクトルの座標位置による微分値とその座標微分動きベクトル空間 (以降, これを単に「特徴空間」と記す) での分布特性を利用する . カメラモーションのみを反映したインライヤは, 各々のカメラモーションに特徴的な分布特性を持つため, それ以外のデータ (アウトライヤ) をある程度排除することができる . よって, アウトライヤの影響をうけにくいという性質がある . さらに, 本アルゴリズムでは,

- 緩い輝度勾配のマクロブロックを除去する (輝度勾配対策)
- 特徴空間における複数のクラスタを用いて GM を検定する (クラスタ選択)

という二つの処理を行なう . これによって, 従来技術におけるボトムアップ的アプローチの問題点である輝度勾配の対策とアウトライヤの影響を少なくできるアルゴリズムを達成できる . ま

¹アフィン変換, ヘルムート変換は平面の回転, 拡大, 縮小を表す変換である .

た，クラスタリングの結果の複数の候補クラスタの中から最適なものを選択するので，クラスタリングの手法によらないアルゴリズム構成になっている．なお，初期動きベクトルは平行移動モデル（2パラメータ）によってマクロブロック単位に求められている．

前章の図 3.2 に局所的な動きベクトルの特徴空間での分布の様子を示す．図 3.2-a は拡大縮小パラメータの分布を示したものである．すなわち，動きベクトルの成分 u, v をそれぞれ x, y で偏微分した値を各偏微分係数を軸とする特徴空間にマップしたものである．同様に，図 3.2-b は回転パラメータの分布を示したものである．すなわち，動きベクトルの成分 u, v をそれぞれ y, x で偏微分した値を各偏微分係数を軸とする特徴空間にマップしたものである．図 3.2-c は各動きベクトル成分を軸とする特徴空間に動きベクトル値をマップしたものである．このように，局所的な動きベクトルを偏微分すれば，GM はあるクラスタに集中する．この性質を利用して GM を算出する．

4.2.1 緩い輝度勾配マクロブロックの除去

勾配の緩やかなマクロブロックは動きベクトルを正確に求められないことが想定される．例えば，アニメーションの画像のように，輝度勾配が小さい場合には動きベクトルの向きがランダムになる．そこで，これらのマクロブロックをその後の処理対象から外す．輝度勾配の判定には対象フレームにおけるマクロブロック内の輝度画素の分散値によって行なう．ある閾値 $TH_{deviation}^2$ 以上の分散値を持つマクロブロックのみをその後の処理対象とする．

4.2.2 拡大縮小パラメータの算出

マクロブロック毎に算出された動きベクトルを式 (4.2) のように偏微分し，2次元特徴空間に射影する（図 3.3-a 参照）．

$$a = \frac{\delta u}{\delta x} = \frac{\delta v}{\delta y} \quad (4.2)$$

但し，近傍ブロックでは特徴空間の座標 $(0, 0)$ にピークが得られる．そこで， n ブロック離れた全てのマクロブロックの組み合わせに対して動きベクトルの差分を求め，これをブロック数 n で正規化して単位ブロックあたりの差分とする．拡大縮小パラメータは原点を通る傾き 45° の直線上に分布するので，この直線から距離 d_{scale} に分布するサンプルの重心を拡大縮小パラメータ値とする．

4.2.3 回転パラメータの算出

同様に，マクロブロック毎に算出された動きベクトルを式 (4.3) のように偏微分し，2次元特徴空間に射影する（図 3.3-b 参照）．

²本章の実験では $TH_{deviation}=50,100$ 等を使用している．

$$b = \frac{\delta u}{\delta y} = -\frac{\delta v}{\delta x} \quad (4.3)$$

実際には n ブロック離れた全てのマクロブロックの組み合わせに対して動きベクトルの差分を求め、これをブロック数 n で正規化して単位ブロックあたりの差分とする。回転パラメータは原点を通る傾きマイナス 45° の直線上に分布するので、この直線から距離 d_{rotate} に分布するサンプルの重心を回転パラメータ値とする。

4.2.4 移動パラメータの算出

式 (4.1) は、以下の式 (4.4) のように変形できる。

$$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} - \begin{bmatrix} a' & b \\ -b & a' \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (4.4)$$

図 3.2 の特徴空間 a, b より算出された拡大縮小パラメータ、回転パラメータを代入し、移動パラメータのみからなる動きベクトル（移動ベクトル）をマクロブロック毎に再計算する。さらに、これらの移動ベクトルを図 3.2 の特徴空間 c にマップし、クラスタのセントロイド（複数）を算出する。

通常は特徴空間での移動ベクトルの重心を算出し、アウトライヤも含めた画像全体で移動ベクトルと重心の 2 乗誤差が最少となる値を求める。しかし、本節ではアウトライヤの影響を排除する新たな手法を検討する。

図 4.2 にクラスタ選択法の概念図を示す。背景に動物体（Moving object）のある風景をパニング等のカメラ操作で撮像した動画を想定する。この動画はノイズなどの影響の無い、理想的な動画であるとする。この画像において、参照画像（Reference image）を用いて背景のみを反映している GM（背景 GM）によって予測画像（Predicted image）を作成する。次に、この予測画像と原画像（Original image）の差分画像（Difference image）を算出する。背景 GM が背景のみを反映しているので、差分の現れる場所な動物体のところのみである。ここで、仮に動物体の面積が他の背景部分の面積よりも小さいとする。背景 GM を含むいくつかの GM の候補があるとすると、この差分の面積が最小になるのは、背景 GM が選択された場合である。

ノイズなどの影響がある一般的な動画においても同様の原理が成り立つ。背景画素数は個々の物体の画素数よりも大きいと仮定する³。GM がアウトライヤの影響を大きく受けているとすると、半数以上を占める背景画素（インライヤ）の予測が外れることになる。ここで、完全にアウトライヤを排除して求められる理想の GM (G_i とする) を仮定する。

G_i を用いて作成された予測画像と原画像の差分の絶対値を求めたとき、差分値があるしきい値 TH より小さくなる画素の数 N_i とする。一方で、アウトライヤの影響を受けた GM を G_o とする。同様に予測画像と原画像の差分の絶対値を求めた時、差分の絶対値があるしきい値 TH よ

³この仮定により、本アルゴリズムは人物の顔がアップになっているような場合など、動物体の画素数が画像の全画素数の反数以上を占める場合には適用できない。

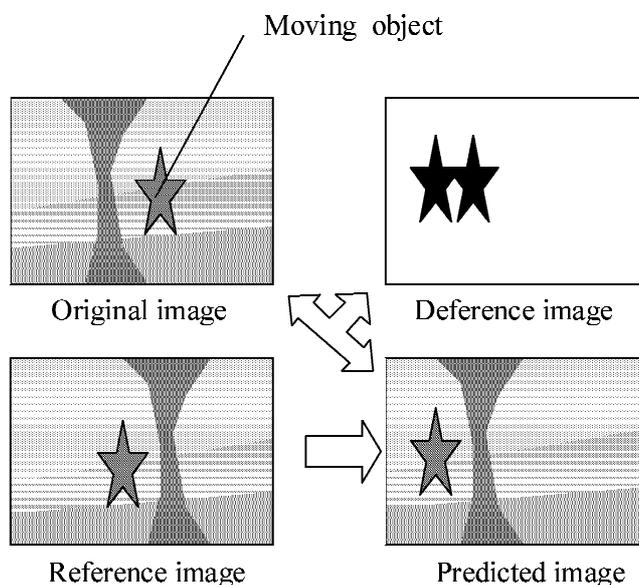


図 4.2: クラスタ選択法の概念 .

りも小さくなる画素の数を N_o とする．アウトライヤの画素数がインライヤの画素数よりも小さいならば必ず $N_o < N_i$ となる．即ち，候補となるクラスタの中でこの N が最大となるクラスタが最もアウトライヤの影響が少ないと考えられる．この性質を利用して，最適な移動パラメータを抽出する．

図 4.3 にクラスタ選択法を用いた移動パラメータ算出の詳細を示す．具体的には，式 (4.4) で求められたマクロブロック毎の移動ベクトルを 2 次元の特徴空間内でクラスタリングする．そこで求められた N 個のセントロイドに対して以下の処理を施す．まず，選択されたセントロイド (仮の GM) を用いて予測画像を算出する．算出された予測画像と原画像の絶対差分値の画像を求める．この差分画像に対してあるしきい値 TH よりも小さい値を持つ画素数 (val) をカウントする．このカウントされた画素数が一番大きいセントロイドがアウトライヤの影響を最も受けにくいものであると考えられる．

4.3 改良アルゴリズムによる GM 算出実験

提案アルゴリズムを表 4.1 に示す 3 つの画像に適用して GM 算出実験を行なった．”MIT sequence”，”Stefan” は MPEG-4 で用いられる標準画像，“Anime” は著者が任意で用意したアニメーション画像である．各々の画像に対して，以下の 3 つの条件の下での GM を算出した．

- 輝度勾配対策，クラスタ選択共に行なわない

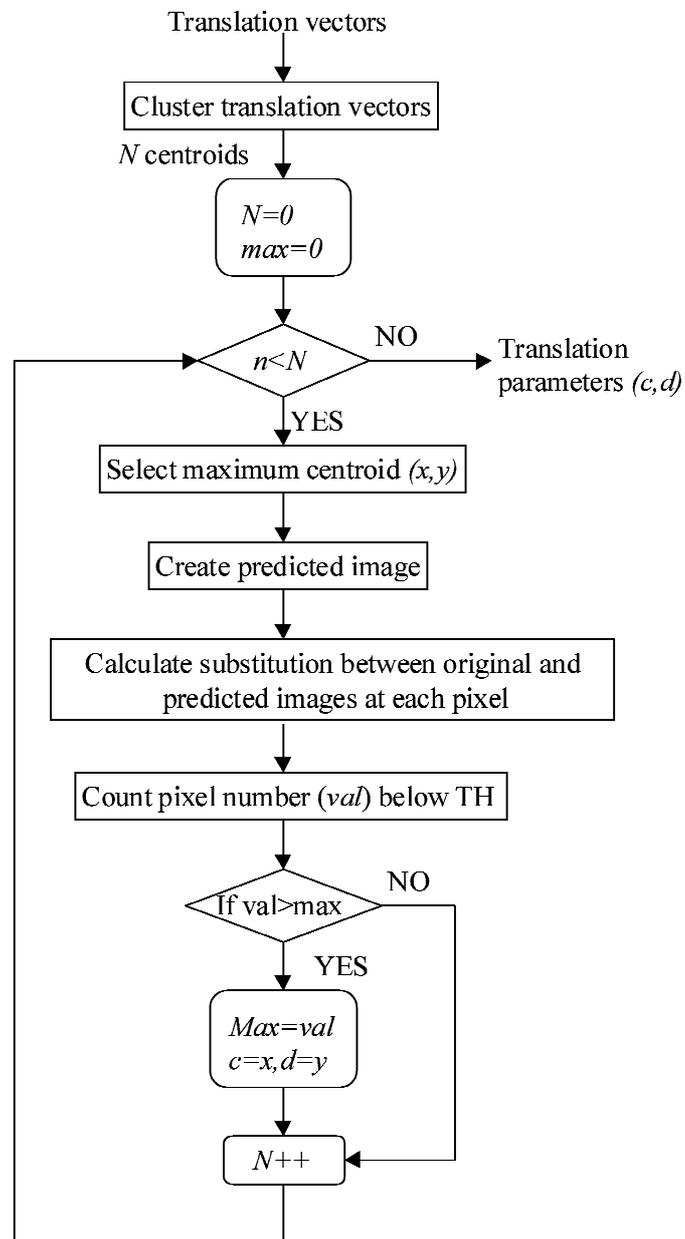


図 4.3: クラスタ選択法の詳細 .

表 4.1: GM 算出実験に用いた画像 .

Video	Type	Frames	Segmentation mask	
			GM	MSE
MIT sequence	SIF	1-150	OFF	OFF
Stefan	SIF	1-150	OFF	ON
Anime	SIF	1-150	OFF	OFF

- 輝度勾配対策のみ行なう
- 輝度勾配対策，クラスタ選択共に行なう

提案アルゴリズムのアウトライヤ除去効果を確認するために，GM 算出に関してはセグメンテーションマスクを用いずに画像全体の情報を使用した．また，クラスタリングにはK 平均法を用いた．

4.3.1 改良アルゴリズムの評価法

改良アルゴリズムを評価するために以下の2つ手法を用いた．

1. 原画像と予測画像から得られる MSE (Mean Square Error) を用いた評価
2. スタティックスプライト生成による主観的评价

前者は，3つの画像において算出した．”Stefan” においてはセグメンテーションマスクを用いて背景部分のみだけで評価を行ない，”MIT sequence”，”Anime” ではセグメンテーションマスクが無いので，画像全体の MSE を算出して評価した．後者は，上記の”Anime”を用いて，第1フレームを基準フレームとして前小節の条件による GM を算出し，さらにスタティックスプライトを生成し，その主観画質から GM の算出精度を評価した．

なお，スタティックスプライト生成には以下の手順を用いた．

1. 隣接フレーム間の GM を算出する
2. 隣接フレーム間 GM を順次，基準フレームからの GM(絶対 GM) に変換する
3. 絶対 GM を用いて，各フレームを基準フレームの座標 (基準座標) にマップする (アライメント) . その際，あらかじめ基準フレームの格子点にあたる画素値を線形補間して求めておく .
4. 基準座標の各格子点において，時間方向にメディアン値を計算し，その値をスタティックスプライトの格子点の値とする .

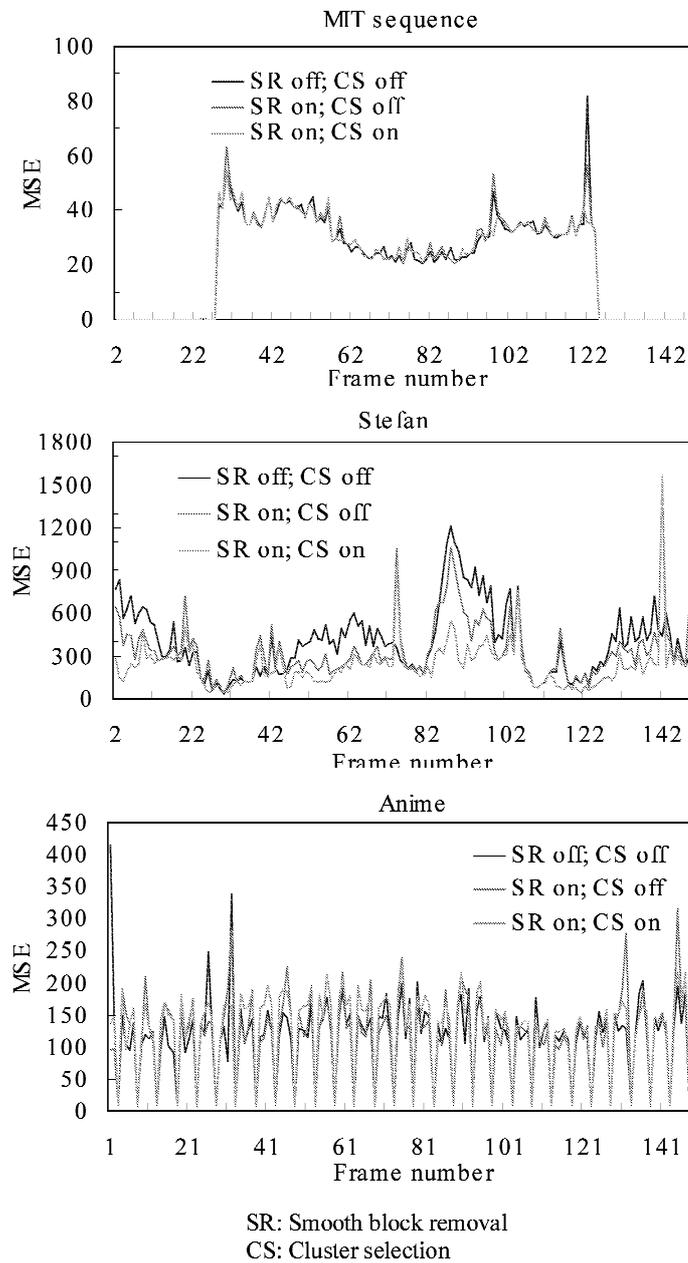
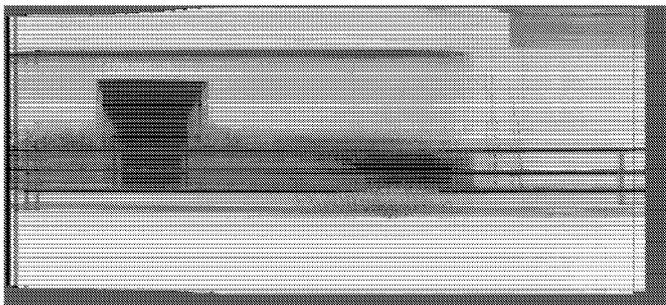


図 4.4: GM 算出実験による提案手法の効果 .

Smooth block removal: OFF, Cluster selection: OFF



Smooth block removal: ON, Cluster selection: OFF



Smooth block removal: ON, Cluster selection: ON

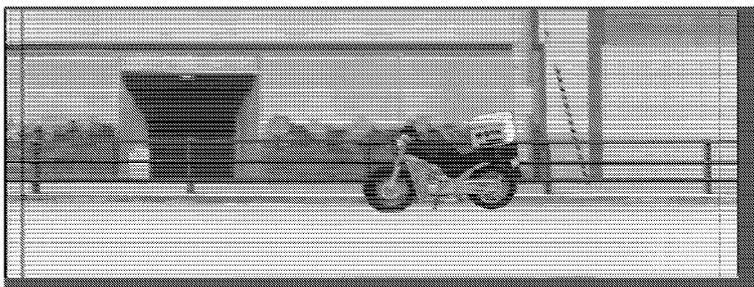


図 4.5: スプライト生成による GM 算出の評価 .

4.3.2 実験結果および考察

輝度勾配対策，クラスタ選択を施した予測画像と着目画像との間の MSE を図 4.4 に示す．全ての画像で輝度勾配対策を施した場合の予測効率が良くなっていることがわかる．特に平坦な色使いを特徴とするアニメーション画像の”Anime”において大きな効果が見られた．一般的に，アニメーションを作成するときに全く同じフレームを複数フレームに使用する場合がある．図 4.4”Anime”において，より MSE が 0 に近くなるフレームが周期的に現れている様子が見られる．すなわち全く同じフレームを複製して次のフレームとしている様子が分かる．

”MIT sequence”，”Stefan”において，MSE の値を 3 つの条件にて比較してみると，輝度勾配対策 ON，クラスタ選択 ON の場合が一番小さな MSE を示している．ここで，”MIT sequence”は元々 1 枚の静止画をパン・ズームのカメラ操作で撮像したものである．すなわち，アウトライヤを含まない背景のみの画像である．同様に，”Stefan”においては，MSE による評価は，セグメンテーションマスクを用いて背景のみを切り出したものを用いている．すなわち，この 2 つの画像においては背景部分のみの MSE で評価をしている．クラスタ選択を行なうと背景部分の MSE が小さくなることから，クラスタ選択を行なった場合の方が行なわない場合よりも正確に背景部分のみを反映した GM を算出できていることがわかる．

一方，”Anime”は，輝度勾配対策 ON，クラスタ選択 OFF の場合の MSE がもっとも低い値を示している．ちなみに，”Anime”はアウトライヤのある画像であるのにも関わらず，セグメンテーションマスクが無いため，フレーム全体で MSE を求めて評価している．ここで，図 4.5 における”Anime”を用いたスプライト作成による GM 算出精度の様子を見ると，輝度勾配対策，クラスタ選択を行なったものがカメラモーションを反映した GM を算出できることが分かる．すなわち，クラスタ選択を行わないで算出される初期重心は，アウトライヤも含めた 2 乗誤差を最少にする値であるので画像全体の MSE は小さくなるが，カメラモーションを正確には反映していないといえる．言い替えれば，クラスタ選択を行うと，フレーム全体の MSE は大きくなる可能性があるが，算出される GM はアウトライヤの影響を含まないものであるといえる．また，図 4.5 より，カメラモーションを正確に反映した GM がスプライトの主観的画品質向上に大きく貢献していることが分かる．

4.4 画像符号化への適用

本章では，MPEG-4 VM Ver.11[48]に基づき，表 4.2 に示す”MITsequence”，”Stefan”，”Coast guard”の 3 種類の画像に対して符号化実験を行なった．各々の画像に対して QP 値を変化させ，(1)LME 符号化，(2)GME 符号化，(3)ダイナミックスプライト符号化，(4)スタティックスプライト符号化を施し，QP 値を変化させて符号量と画品質を比較検討した．3 つの画像とも，MPEG-4 で使用されている標準画像である．なお，GM 算出はセグメンテーションマスクは使用せず，フレーム全体の情報を用いた．

表 4.2: 画像符号化実験に用いた画像 .

Video	Type	Frames	Framerate
MIT sequence	SIF	1-150	30
Stefan	SIF	1-150	30
Coast guard	SIF	1-150	30

4.4.1 符号化方法

以下に、各符号化方法を簡単に説明する .

LME(Local motion estimation) 符号化

DCT + 動き補償による通常の符号化 . 動き予測はマクロブロック単位の平行移動モデルを用いる . DCT 係数 , 動きベクトル , 動き補償による差分情報が符号化される .

GME 符号化

GM 予測は提案アルゴリズムで算出されたものを用いる . 全フレームにおいて GME=ON とし , GM と GM を用いた動き補償による原画像との差分を符号化する .

ダイナミックスプライト符号化 (DS)

提案アルゴリズムによる GM を用いて時間的に変化するダイナミ”Anime”において , ックスプライトを生成する . GM とスプライトから切り出された予測画像と原画像の差分を符号化する . 基本的に GME 符号化と同じであるが , ダイナミックスプライト符号化ではデコーダ側にスプライトを格納するための大きなバッファを確保する必要がある .

スタティックスプライト符号化 (SS)

提案アルゴリズムによる GM を用いてビデオシーケンス全体の静的 (スタティック) スプライトを生成する . GM とスプライトを符号化する . 原画像との差分は一切符号化しない .

4.4.2 実験結果および考察

表 4.3 に使用した画像 , 主な符号化パラメータならびに符号化結果を示す . GM 算出に用いた平行移動モデルによる動きベクトルはハーフペルの検出精度のものである . 算出された実数値の

表 4.3: 符号化効率ならびに画品質 .

Image	QP	Coding efficiency(upper)[kbps] and SNR(lower)[dB]				
		LME	GME	DS	SS (key=1)	SS (key=75)
MIT sequence	1	— —	— —	— —	98.92 26.51	285.48 24.28
	7	— —	— —	— —	27.76 26.03	70.55 24.42
	12	385.59 31.18	334.59 31.14	336.45 31.14	18.49 25.39	50.09 24.34
	21	— —	— —	— —	12.45 24.43	36.19 24.14
	31	188.61 26.63	96.32 26.36	99.13 26.38	9.51 23.50	29.52 23.91
Stefan	1	— —	— —	— —	163.01 19.19	135.33 19.25
	12	747.61 29.69	745.27 29.68	742.91 29.67	27.23 25.39	23.04 24.34
	21	— —	— —	— —	17.79 19.00	14.71 19.00
	31	251.75 24.22	229.72 24.20	228.19 24.22	13.43 18.86	10.98 18.82
Coast guard	1	— —	— —	— —	86.66 20.74	86.86 20.94
	12	472.21 29.96	459.40 29.91	458.92 29.91	11.08 20.78	11.17 20.91
	21	— —	— —	— —	7.84 7.86	7.86 20.86
	31	180.45 26.32	142.05 26.16	143.20 26.18	6.56 20.71	6.53 20.80

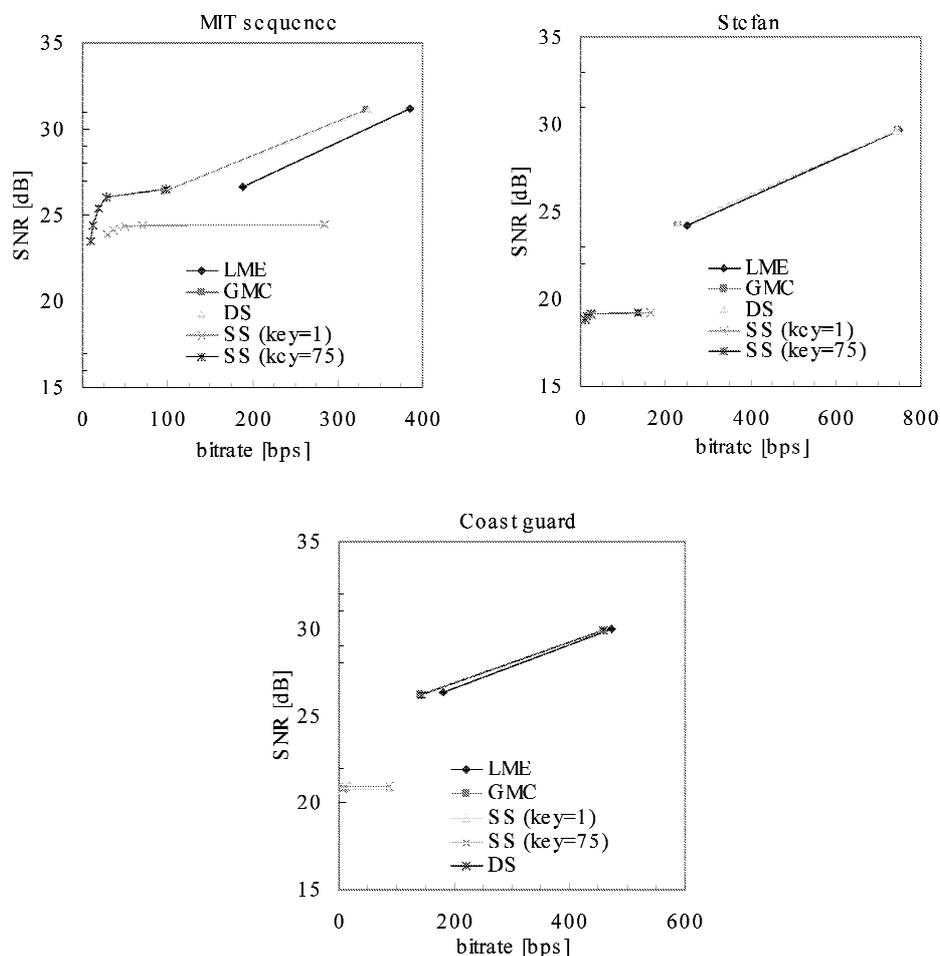


図 4.6: 復号画像の原画像に対する SN 比の変化 .

GM は文献 [3][48] にのっとり、4 点の座標変換として符号化される。⁴図 4.6 に符号化効率と画品質 (SNR) の関係を示す。図 4.7 に生成されたスプライトの例を示す。図 4.8 に”MIT sequence” による符号化して復号した画像の例を示す。ここで、スタティックスプライトの符号量は、他の符号化方式と比較するために、便宜上、シーケンス全体のビット数を秒で除した値を用いている。図 4.7 より、スプライトが良好に生成されているのが分かる。フレームの統合時に時間方向にメディアン値をとっているため、動物体 (テニスプレーヤ、船舶など) が消去されている。図 4.6 より、LME、GME、DS の 3 者と SS の符号化の特性の相違が明確に表れていることが分かる。前者は高いレートで高い SNR を達成できるが、QP を最大値の 31 に設定しても LME では約 200kbps

⁴一度算出された実数値の GM を元の平行移動動きベクトルの検出精度に合わせて丸める処理は行っていない。よって、GM の精度は最初に算出される平行移動モデルによる動きベクトルの検出精度によらない結果になっている。

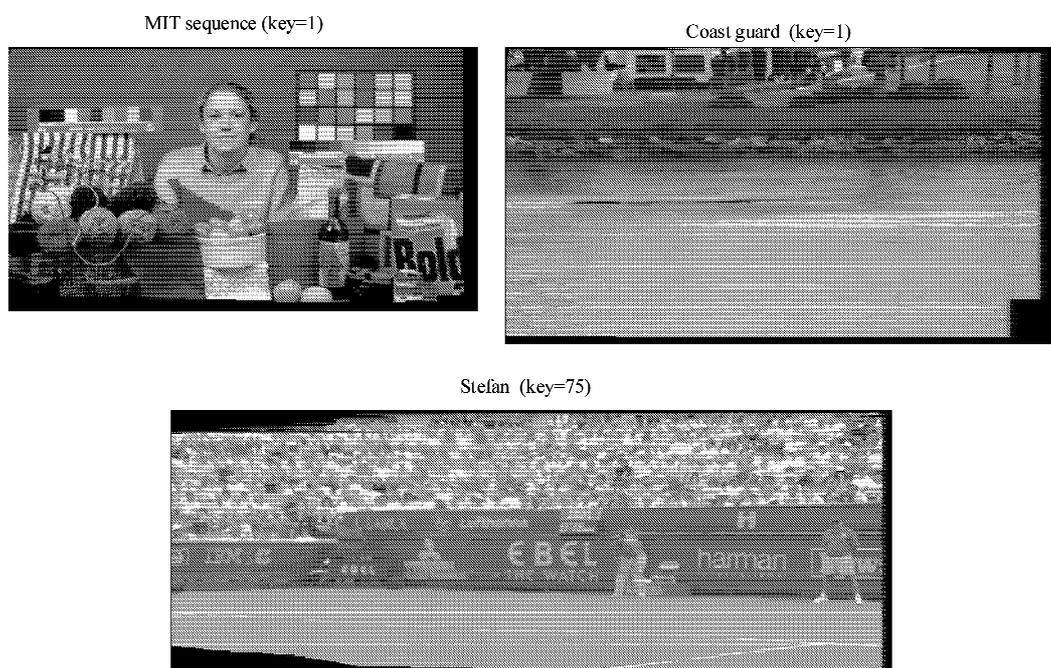


図 4.7: 生成されたスタティックスプライト。

が最高，GME，DS では”MIT sequence” が約 100kbps と低レート化には限界がある．また，3 者のうち GME と DS はほぼ同一の特性を示すことがわかる．GME と DS は”MIT sequence” において LME の約 $3/4$ の符号量を達成しているが，符号化効率の大幅改善にはいたらない．一方，SS においては，100kbps ～ 数 kbps の超低レートを達成できることが分かる．これは LME の限界符号量の約 $1/2 \sim 1/30$ に相当する．

SNR による画品質を評価すると，逆に SS は低い値となる．これは，SS においてはスタティックスプライト生成時のアラインメント，インテグレーション処理の時に，線形補間されて惚けた画素が選択される，あるいは，GM の若干の誤差の蓄積による画素のずれなどが原因と考えられる．主観画質に着目して，図 4.8 を参照し，符号量と主観画質で評価を行なう．画像は”MIT sequence” の第 30 フレームを用いた．上段には LME，GME，DS における QP = 12 の画像，中下段には基準フレームを 1 フレームおよび 75 フレームにおいた SS の QP=12,21,31，すなわち低符号量時の復号画像を示した．SS(Key=75) の QP=12 および 21 が LME QP = 12 とほぼ同じ主観画質を与えている．符号量に着目すると，LME と比較して前者が約 $1/8$ ，後者が約 $1/12$ の符号量を達成している．

一方，”Stefan” や”Coast guard” に対しては，元々，観客，テニスプレーヤ，船舶，波などのアウトライヤが含まれた画像であり，ここから忠実に背景を抽出するのは困難である．この様に背景が時間的に変化する場合はスタティックスプライトの精度向上手法として，文献 [3] で提案されている Low-latency sprite があげられる．これは，最初は精度の粗いスタティックスプライ

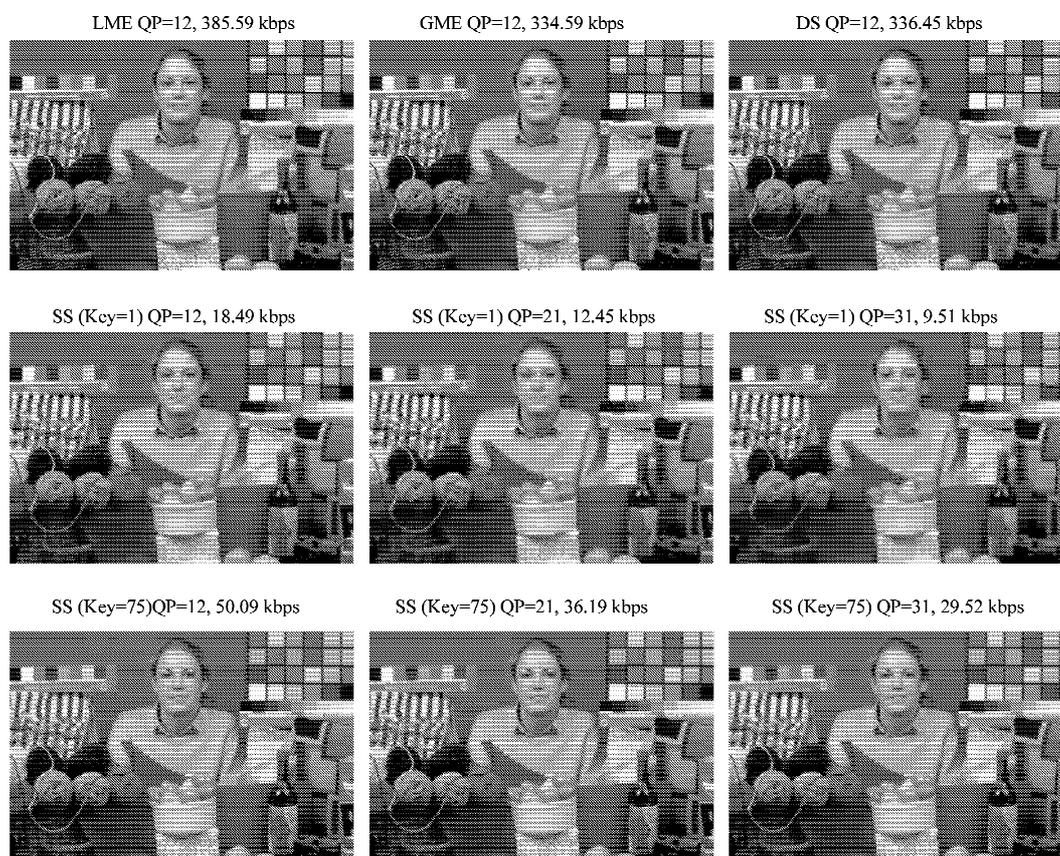


図 4.8: 復号画像の例 .

トを送信して、徐々に精度を上げていく手法である .

スタティックスプライトの符号量の大部分をスプライトデータが占めているため、オフラインで予めスプライトを送信した場合、伝送する情報はGMのみとなるので、さらに高い符号化効率を達成できる . 上記の”MIT sequence” の場合、150 フレーム分のGMのビットレートは約 5kbps であり、LME 符号化の全ビットレートの数十分の 1 以下である . また、”MITsequence” のスタティックスプライトの符号化結果から分かるように、10kbps でフレームレート 30 を達成出来る . このように、スタティックスプライト符号化は低ビットレートで高フレームレートを実現する場合に有効である . ただし、SS は以下の二つのデメリットがあることを明記しておく . 一つは、他の 3 手法が One-pass 符号化であるのに対して SS は Two-pass 符号化であるためにアプリケーションが限られてくることである . もう一つは、SS は差分を符号化しないために、場合によっては視覚的におかしな画像が生成される場合があり、適用できる画像に制限があることである . また、スプライト生成の精度が画品質に大きく影響するため、今後はスプライト生成の高精度化が研究の課題である .

4.5 本章のまとめ

本章では、輝度勾配対策とクラスタ選択を特徴としたアウトライヤの影響の少ないグローバルモーション算出アルゴリズムを提案し、算出精度の評価を行なった。輝度勾配対策は、アニメーション画像において効果を発揮した。クラスタ選択はカメラの動きのみを反映したグローバルモーション算出においてその精度向上に大きく寄与することが分かった。また、上記の背景画像に対して、MPEG-4 Ver.11における(1)LME符号化、(2)GME符号化、(3)ダイナミックスプライト符号化、(4)スタティックスプライト符号化を行った。現状では再現の忠実性に課題を残すものの、スタティックスプライトはその他の符号化方法に比べて高い圧縮効率を達成出来ることが分かった。今後はスプライト生成の高精度化、Low-latency スプライトの検討が課題である。

第5章 2層ビデオオブジェクト自動生成とスプライト符号化への適用

5.1 はじめに

筆者は、前章までに、インターネットやモバイル通信環境に適用できる超低ビットレート符号化方法の研究開発 [49][50][51][52][53] を行ってきた。従来符号化法 (H.26X 等) にて、超低レート (128kbps 以下) 等で画像を符号化すると、十分なフレームレートや画質を与えられないという問題があった。これは MC+DCT ベースのフレーム毎の符号化方式の限界であり、これに変わる圧縮法が望まれていた。

本章では、その実現方法の一つとして、新しい画像符号化標準 MPEG-4[3] における「スプライト符号化」に着目し、動画像を前景、背景の2層のビデオオブジェクトに自動的に切り分け、それぞれを MPEG-4 オブジェクト符号化、MPEG-4 スプライト符号化する手法を提案する。

MPEG-4 では、オブジェクトベースの符号化を取り入れている。オブジェクト符号化では画像を意味あるオブジェクト単位に分割し、そのオブジェクト毎に符号化し、多重化して伝送する。利点として、オブジェクト毎にフレームレートや画質を変えることが可能である。スプライト符号化はオブジェクト符号化の一形態であり、スプライトを一つのオブジェクトとして扱うことが可能である。スプライト符号化は、カメラの動きのみを反映した背景画像に対して、カメラモーションに基づき画像を貼り合せて一枚のパノラマ画像 (背景スプライト) を作り、これを符号化に利用する¹。大量のフレームが一枚の静止画像で表現でき、原画像との差分を一切符号化しないことから従来法に比べて劇的な符号量削減が期待できる [9][11][35][51][54]。

しかし、スプライト符号化およびオブジェクト符号化の課題の一つとして、オブジェクトの自動生成が困難であるという点がある。例えば、動物体の抽出法においては、半自動の抽出法 [20]、自動抽出法 [27][28][32][55]、自動追跡法 [56] が提案されているが、オブジェクトが唯一存在する場合に限られる。多数のオブジェクトが存在する場合の異なるフレーム間での同一オブジェクトの対応問題が検討されされていないため、任意の画像に適用することが困難であった。

筆者は文献 [50][51] において、背景スプライトの自動生成法を提案し、生成された背景スプライト部分において、MPEG-4 スプライト符号化し、従来符号化法と比較して数倍の圧縮効率を達成している。しかし、前景部分の抽出ならびに前景と背景をあわせた実際の符号化については未検討であった。本論文では、動画像を超低ビットレートでの符号化する方式の確立を目的とし、文献 [50][51] の課題を克服するために、動画像から背景スプライトと前景オブジェクトを自動生

¹背景画像以外にも、オブジェクトの大局的な動きが一組の変換パラメータで表される場合はスプライトを構築することが出来る。背景スプライトはその一例である。

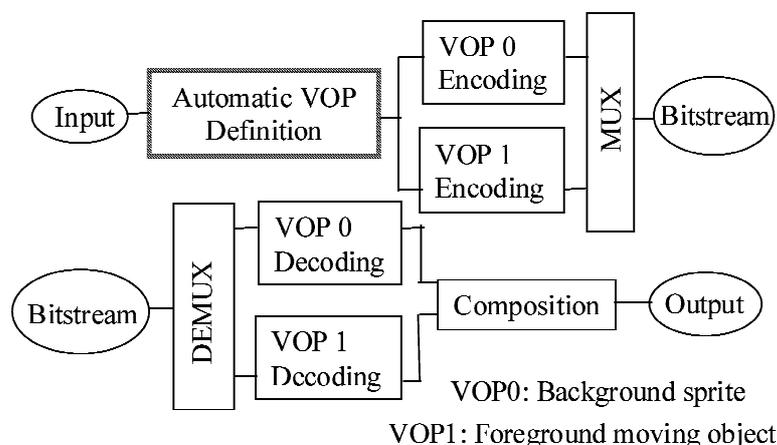


図 5.1: MPEG-4 符号化復号モデルと 2 層ビデオオブジェクト。

成するアルゴリズムを提案する。また，提案アルゴリズムによって生成した前景，背景のビデオオブジェクトをそれぞれ MPEG-4 オブジェクト符号化ならびにスプライト符号化して，スプライトを用いない MPEG-4 Simple profile の符号化（ノーマルモード）²との符号量の比較検討を行う。

はじめに，オブジェクトの対応問題を回避するために，前景オブジェクトと背景スプライトの 2 層からなる VOP(Video Object Plane) を定義する。VOP は従来符号化法のフレームに相当するもので，ビデオオブジェクトは一連の VOP の集合体である。2 層ビデオオブジェクト生成アルゴリズムは，スプライト生成に特化したグローバルモーション（GM）算出法，高品質背景スプライト生成法，前景オブジェクト抽出法からなる。GM 算出法には筆者らが文献 [50][51] で提案しているカメラモーション算出法を用いた。高品質スプライト生成においては，本章では，従来の時間メディアン法 [10]，オーバーライト法³の長所を生かした方法を新たに提案する。本手法では，前景のない，高品質なスプライト生成することが可能である。前景オブジェクト抽出法は，原画像と背景スプライトの差分を基本とし，GM のずれに対してロバストな方法を提案する。自動生成における前景オブジェクト形状は輪郭が複雑であること，孤立点が多数含まれることなどから予測がほとんどあたらず，その結果イントラ符号化されて符号量が増大する。この問題を回避するために形状をマクロブロック⁴で近似する方法を提案する。さらに，本章では，提案アルゴリズムを用いて生成した二層ビデオオブジェクトを MPEG-4 スプライト符号化し，VOP 構造を持たない MPEG-4 通常符号化方法と比較検討を行った。特にフレームレートや前景の大きさが符号量に与える影響について調査し，スプライト符号化が劇的な圧縮率を達成できる条件について考察を行った。さらに，同条件下にて同一ビットレートでの符号化実験を行い，フレームレ

²以降，スプライトを使用する MPEG-4 符号化をスプライトモードでの符号化，スプライトやオブジェクトを使用しない従来方式を踏襲した MPEG-4 符号化をノーマルモードによる符号化と呼ぶことにする。

³画像を上書きしながら貼り合わせてパノラマ画像をつくる一般的な手法。

⁴16 画素四方で形成される，処理単位。

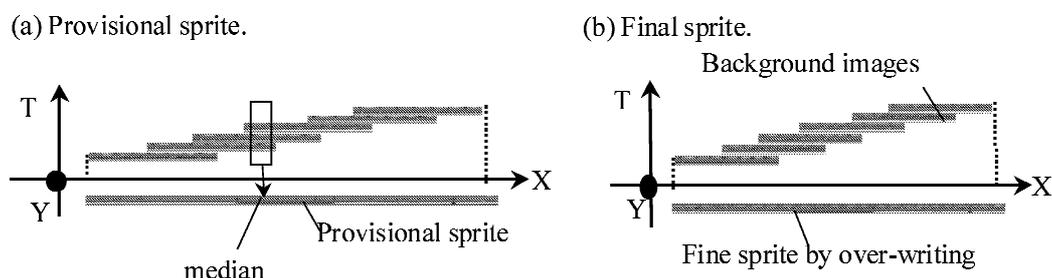


図 5.3: XYT 時空間における画像のアラインメントの様子。(a) 仮スプライト。(b) 本スプライト。

時間方向にメディアン（中央値）をとってスプライトの値として、動物体を消去した仮スプライトを生成する。さらに、仮スプライトと原画像の差分画像から前景候補画像と背景画像を生成する。前景候補画像はマクロブロック近似されて最終的な前景画像となる。一方で、原画像から前景領域を除かれた背景画像を幾何変換して基準座標上にマップして背景スプライトが生成される。以下にその詳細を述べる。

5.2.1 GM 算出

スプライトを生成するためには、カメラ操作を反映した GM を算出する必要がある。本章では、文献 [50][51] において筆者らの提案した GM 算出法を用いる。本方法は以下の原理を利用している。

1. カメラモーションを反映した局所的動きベクトルにおいて、その x 方向（水平方向）ならびに y 方向（垂直方向）の変化量を軸とする特徴空間で独特の分布を示すことを利用
2. カメラモーションの候補となる複数のクラスタをそれぞれ検定し、最適なものを GM として算出（クラスタ検定）

算出された各々のフレーム間 GM を用いて、基準座標からの GM にそれぞれ変換される。以降、GM とは基準座標からの GM を指す。

5.2.2 背景スプライト生成

前景の無い、かつ、高品質なスプライトを生成するために、本章では従来の時間メディアン法 [10]、上書き法の長所を利用した方法を提案する。生成されるスプライトは仮スプライトと本スプライトの二種類である。

図 5.3 に画像が時間方向へ統合される様子を示す。スプライトとは画像を模様がつながるように並べてパノラマ画像にしたものである。前節で算出された GM を用いて各々の原画像を幾何変形し (x, y, t) 基準座標系の格子点上にマッピングする。基準座標系の任意の座標 (x, y) に注目す

ると、時間 (t) 方向に複数の画素が重なっている。これら画素値のメディアンをとってその座標の値とする。ここで、動物体領域がその座標にマップされた画素数の半分より少ない場合を想定する。この場合は、座標上には背景を代表する画素が半数以上を占めるので、背景を代表する画素がメディアンとして選択される。これをすべての座標で行うと動物体が消去されたスプライトが生成される。本論文では、このスプライトを仮スプライトと呼ぶ。しかし、時間メディアン法によって生成された仮スプライトは、動物体は消去されているものの、微小な GM のずれによって座標上に線形近似された値が候補として並んでいることから、スプライトが若干ぼけるという性質があった。そこで、この仮スプライトを前景抽出に利用する。

一方、上書き法は、画素をそのまま基準フレームに貼りつけていくので画質の良いスプライトが生成できるが、一番手前に来るフレームの前景と、各フレームの端の前景オブジェクトが残るという問題があった。よって、動物体を抽出した後、動物体を取り去った部分を背景画像とし、それを基準座標に GM で幾何変換して順次上書きし、最終的な高画質スプライトを生成する。本論文では、これを本スプライトと呼ぶ。

5.2.3 前景オブジェクトの抽出

前景オブジェクト生成は、前景候補領域の抽出と前景形状のマクロブロック近似の二段階の手順を経て生成される。さらに、前景比率によって抽出される前景を制御する。次節以降にその詳細を述べる。

前景候補領域の抽出

最初に、仮スプライトから GM によって切り出された背景画像と原画像を用いて、絶対差分画像 D を生成する。ここで、絶対差分画像 D を適当な閾値 Th_0 で二値化し、二値化画像 T を得る。差分の大きいところは動物体である可能性が高いので、この二値化画像 T を前景候補領域とする。

マクロブロック単位の前景オブジェクトの特定

自動生成における前景オブジェクト形状は輪郭が複雑であること、孤立点が多数含まれることなどからほとんど予測があたりずイントラ符号化され、その結果、符号量が増大する。本章ではこの問題を回避するために、オブジェクト形状をマクロブロックで近似する方法を提案する。図 5.4 に前景形状のマクロブロック近似の様子を示す。

MPEG-4 において、オブジェクト形状符号化は可逆、非可逆の二通りのモードが用意されている [3][54]。その粗さはパラメータ $alpha_{Th}$ で表される。一番粗い形状表現はマクロブロックで近似される。具体的には、マクロブロック内に前景情報が半数以上あればマクロブロック全体を前景、半数より少なければマクロブロック全体を背景とするものである [54]。元の形状を図 5.4-(a) としたとき、MPEG-4 における形状符号化の一番粗い近似法を図 5.4-(b) に示す。この方法は前景オブジェクトがかなり侵食されて視覚的に妨害になるという問題があった。そこで、本章では、

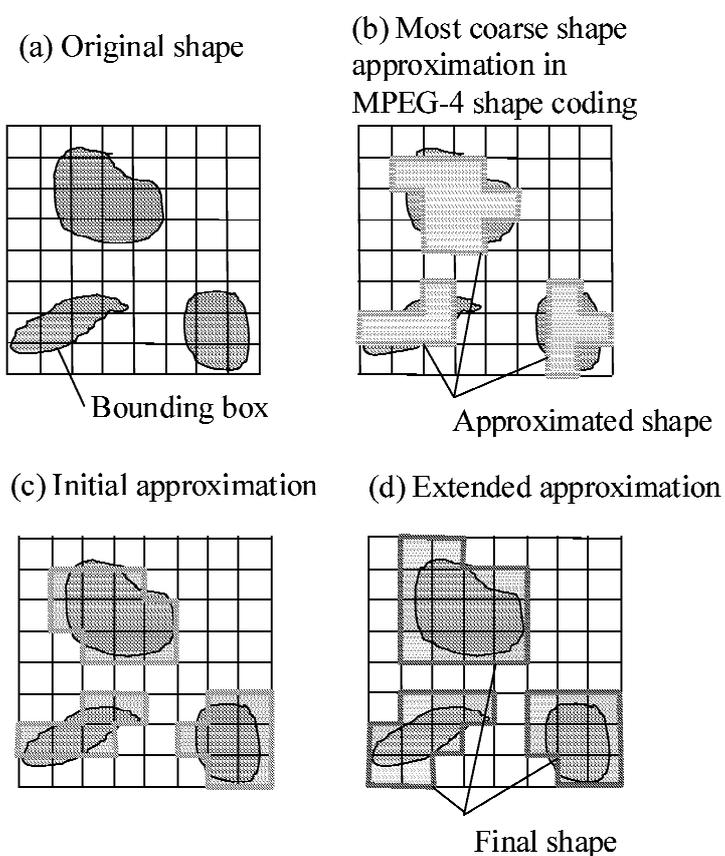


図 5.4: 前景オブジェクト形状のマクロブロック近似の様子。(a) 元の形状, (b) 一番粗い形状近似, (c) 提案手法における初期マクロブロック化, (d) 提案手法における拡張マクロブロック化。

新たなマクロブロック近似法を提案する。本手法は初期マクロブロック近似と拡張マクロブロック近似の二つの手法からなる。図 5.4(c) に初期マクロブロック近似, 図 5.4(d) に拡張マクロブロック近似の様子を示す。マクロブロック内に前景が閾値 Th_1 画素以上ある場合, マクロブロック全体を前景と判断する。それ以外のは背景とする。次に, 初期マクロブロック近似では前景となったマクロブロックに隣接する背景マクロブロックにおいて, そのマクロブロック内に前景画素が閾値 Th_2 ($Th_2 < Th_1$) 画素以上あるものを前景と再判断する。拡張マクロブロック近似は N 回行う。このように, 提案したマクロブロック近似法は, 初期マクロブロック化によって前景オブジェクトの大まかな場所を特定する。さらに拡張マクロブロック化によってその周辺の領域も前景オブジェクトに統合するという作用がある。提案手法でマクロブロック近似されたオブジェクト形状は MPEG-4 形状符号化における可逆モードで符号化される。

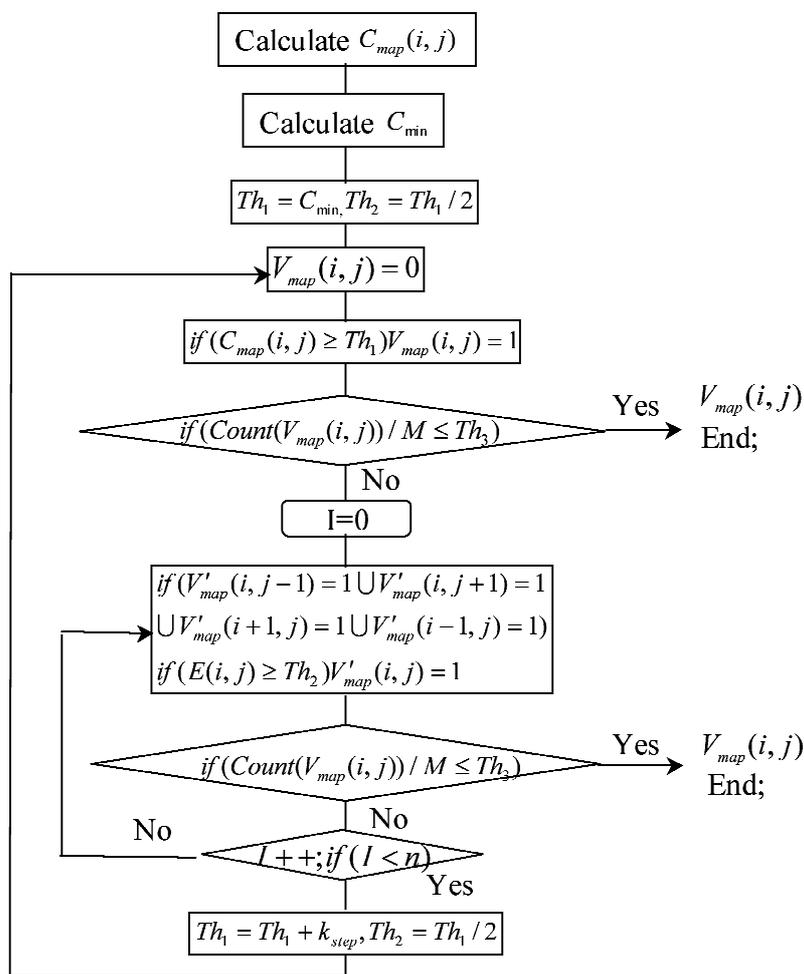


図 5.5: 前景比率固定下における前景オブジェクト抽出アルゴリズム .

前景率固定下における前景オブジェクト抽出

最適な閾値 Th_1, Th_2 は画像に依存する．例えば，複雑な背景をもつ画像では，GM が誤抽出された場合，差分による前景候補領域は必然的に多くなる．これをすべて前景領域と判定すると，前景の面積が大きくなり，符号量の増大につながる．一方で，一様なテクスチャを背景に持つような画像では，GM 誤抽出によって抽出される前景候補領域は少ないと考えられる．そこで，閾値 Th_1, Th_2 を自動で決定するために，一定前景比率の前景オブジェクトを抽出するアルゴリズム（図 5.5）を提案する．基本的には，前述の初期マクロブロック化と拡張マクロブロック化の組み合わせを繰り返し行い，任意の前景比率（前景オブジェクトの大きさの画像全体の大きさに対する割合）を処理の終了判定に使用する．

具体的なアルゴリズムの流れを以下に説明する．最初に， V_{map} を初期化する． V_{map} は，前景

表 5.1: MPEG-4 スプライト符号化実験に用いた画像 .

Video	Type	Frames	Content
Horsrace	SIF	150	pan, two horses and a car
Soccer	SIF	150	pan and tilt, several players
Athlete	SIF	150	pan, a couple of runner
Stefan	SIF	150	pan and zoom, a player
Board	SIF	150	pan and tilt, a border

候補マクロブロックの場合は $V_{map}(i, j)$ ((i, j) はマクロブロック位置を示す) に 1 を格納し, 背景マクロブロックは 0 を格納するものである. 前述の二値化画像 T を用いて, 各マクロブロック中の前景候補画素数をカウントしてマップ $C_{map}(i, j)$ を算出する. さらに前景候補画素数の最小値 C_{min} を算出する. 初期マクロブロック化の閾値 Th_1 を C_{min} , 拡張マクロブロック化の閾値 Th_2 を Th_1 より小さい任意の値に初期設定する. ここでは $Th_2 = Th_1/2$ を用いている. 初期設定された閾値 Th_2 と Th_1 を用いてマクロブロック化を行う. ここで前景比率 ($V_{map}(i, j) = 1$ のマクロブロック数を総マクロブロック数 M で除した値) を計算し, 任意の割合 Th_3 を下回らない場合, 処理を続行する. この時, マクロブロック化閾値 Th_1 と Th_2 を $Th_1 = Th_1 + k_{step}$, $Th_2 = Th_1/2$ のように更新する. ここで, k_{step} は任意の刻み幅である. 次節に述べる実験では $k_{step} = 2$ を用いた. 前景比率が任意の割合 Th_3 を下回った時, 処理を終了する. 処理を終了したときのマクロブロック化結果が前景オブジェクトとなる.

形状情報を有する標準画像”Stefan”において, 人物部分の形状情報をマクロブロック近似したところ, ロスレスで形状符号化する場合に比べて形状符号量が 1/10 程度に抑えられた. マクロブロック近似の他のメリットとして, ソフトウェアデコーダ実装時の処理コスト軽減があげられる. 形状情報の復号, パディングといった処理をしないのでリアルタイム復号が求められるアプリケーションに好都合と考えられる.

5.3 MPEG-4 符号化実験

スプライト符号化は, 複数フレームにわたる背景部分を一枚の静止画で表し, 原画像との差分を一切符号化しないことから, 従来符号化に比べて劇的な圧縮率を達成可能であることが報告されている [10][35][50][51]. しかし, 画像によってはスプライトモードで符号化しても劇的な符号量削減ができない場合が考えられる. その要因としてフレームレートと前景比率が上げられる. フレームレートが少ない場合, 必然的に背景スプライトで表現できるフレーム数が少なくなり, 劇的圧縮効果が得られない可能性がある. また, 前景比率がある程度以上ではスプライトモードで符号化しても符号量削減にはならないことが予測される. これは, 前景比率が 100% の場合 (画面全体が前景となった場合), ノーマルモードに比べてスプライトモードはスプライトの分だけ

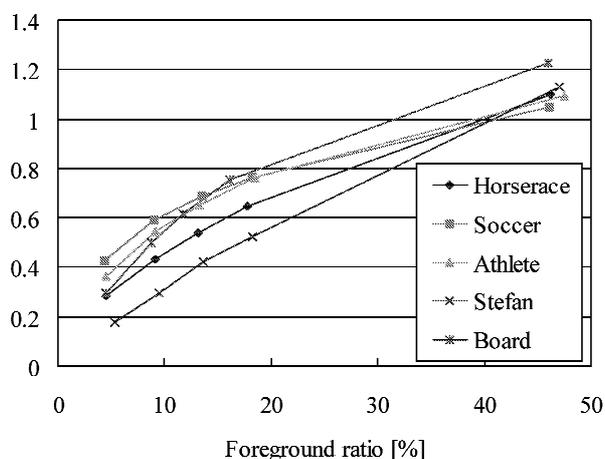


図 5.6: 前景比率と MPEG-4 スプライト符号化における符号量の関係．MPEG-4 通常符号化時の符号量との比で表現．

情報量が増大することから明白である．ある程度の前景比率のところではスプライトモードとノーマルモードの符号量が同じになると考えられる．4.1 及び 4.2 節では，表 5.1 に示すカメラ操作を含む 5 つの画像において，提案アルゴリズムによって生成されたビデオオブジェクトをスプライトモードで符号化 (図中では”Sprite”) し，スプライトを用いないノーマルモード (同，図中では”Rectangle”) と比較し，劇的な圧縮が可能である動画像の条件を検討した．ここでは，量子化パラメータ QP を固定にして両モードの符号量を比較検討した．これは，文献 [12] における，人工的に作られた前景領域を持たない画像 (”MITsequence”) に関して，同程度の QP でノーマルモードとスプライトモード⁵で符号化した画像は同程度の主観画質を与えるという知見に基づく．本実験では”Stefan”を除いては背景がカメラに因る動きを除いてはほぼ静止しているものを用いた．よって，人工画像から得た知見を自然画像にほぼ適用できると考える．また，4.3 節では，低レート時 (128kbps, 64kbps) にて両モードでの画質とフレームレートを客観的に評価した．

ここで，ノーマルモードとスプライトモードの前景オブジェクト部分の符号化において，GOV の取り方によって有利不利が生じないように，一つのシーケンスにつき一つの GOV 構造とした．すなわち，最初の I-VOP 以外はすべて P-VOP として符号化される．ノーマルモードにおいては一番符号量の小さくなる GOV 構造である．また両モードにおいて，MV 探索範囲は ± 32 画素とした．スプライトモードにおける GM はこの探索範囲で求められた MV を利用して算出した．

⁵文献 [12] では各々 LME:Local Motion Estimation, SS:Static Sprite と表記．

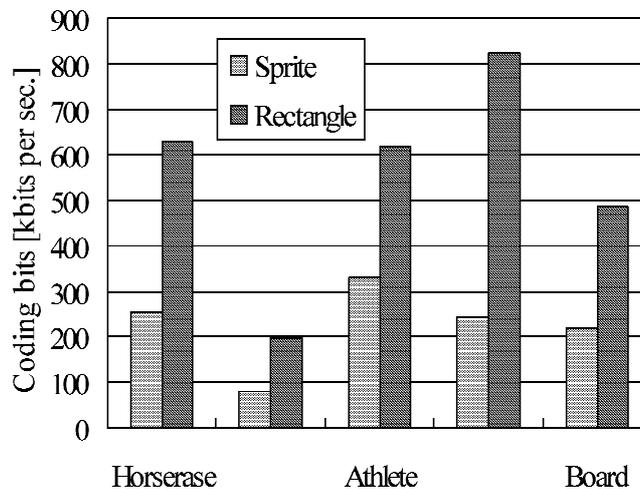


図 5.7: 前景比率 10-15 %程度, 30fps における MPEG-4 スプライト符号化と MPEG-4 通常符号化の符号量の比較 .

5.3.1 前景比率と符号量の関係

最初に, 圧縮率に大きく寄与すると考えられる前景比率を 5-40 % の範囲で可変にして両モードでの符号量測定を行った. 画像サイズは SIF, フレームレートは 30fps (frame per sec.), 量子化パラメータ QP は 12 で固定とした. また, スプライトモードの符号量は, 前景オブジェクトと背景スプライトの符号量の合計を 1 秒あたりの符号量に換算した値を用いた. また, 各オブジェクトの符号量はそれぞれ, テクスチャ, モーション, 形状の符号量の合計である.

図 5.6 に前景比率を変化させ, ノーマルモードとスプライトモードの符号量の比率を測定した結果を示す. 横軸は前景比率, 縦軸にスプライトモードの符号量をノーマルモードの符号量で除した値である. 縦軸が値 1 を取るのは両符号化モードによる符号量が同じであることを示している. 前景比率が 10-15 % 以下の場合, スプライトモードがノーマルモードの半分以下の符号量を達成できた. また, 前景率が 40 % 程度でスプライトモードとノーマルモードの符号量が同程度になった.

5.3.2 フレームレートと符号量の関係

次に, 任意の前景比率において, 五つの画像のフレームレートを 5, 10, 15, 30fps と可変にし, 両モードでの符号量を測定した. ただしスプライトはフレームレート 30fps の状態で作成したものを符号化に適用している. 量子化パラメータは QP=12 で固定とした.

図 5.7 に前景比率 10-15 % 程度の時の五つの画像のスプライトモードとノーマルモードの符号量の関係を示す. スプライトモードがノーマルモードの $1/2$ から $1/4$ の符号量を達成しているこ



図 5.8: $QP = 12$ 固定で符号化時の符号化画像の例 (“Horserace”) . 左: ノーマルモード, 右: スプライトモード .

とがわかる . 図 5.8 に $QP = 12$ で符号化時の “Horserace” の両モードの符号化画像を示す . 主観画質に関してはほとんど差が無いことが分かる . ちなみに画像 “Horserace” のスプライトモードにおける平均 SNR は 26.22[dB] , ノーマルモードにおける平均 SNR は 30.98[dB] であった . SNR から主観画質が単純には比較できないことについては以降の 3.3 節で述べる .

図 5.9 に本アルゴリズムによって生成された背景スプライトの例 , 図 5.10 に原画像 , 前景オブジェクト , 合成画像の例 (符号化画像) を示す . 前景比率は 10-15 % 程度である . 前景の消去されたスプライトが生成されていることがわかる . また , 前景オブジェクトはマクロブロック単位に抽出されるので本来は背景に属する領域も前景オブジェクトとして抽出されている . しかし , 背景スプライトから背景を切りだし , 背景画像と前景を合成すると違和感の無い画像が生成されていることがわかる . 合成部を拡大してみると , $QP = 12$ で符号化時のブロック歪と区別がつかないほど良好に合成されていることがわかる .

また , 図 5.11 にスプライトモード及びノーマルモードにおいて , フレームレートを変化させた場合の符号量を示す . 前景比率は 10-15 % 程度である . 同程度の画質を達成するのに , スプライトモードはノーマルモードに比べて $1/2$ から $1/4$ の符号量で実現できている . この傾向は “Horserace” を除いてどのフレームレートにも当てはまる . “Horserace” は被写体やカメラの動きがかなり速いため , 低フレームレートではノーマルモードでは動きベクトル探索範囲が十分確保できず , 動き補償が効果的でなかったためと推定される . “Athlete” を例にとると , ノーマルモード 200kb/s 相当の符号量では 5fps (frame per sec.) しか達成できないのに対し , スプライトモードでは 20fps 以上を達成可能である . スプライトを使用すると , インターネットのような狭帯域環境下でもたくさんのフレームを表現でき , 低ビットレートでも動きなめらかな映像を SIF 等の大きな画像サイズで実現可能であることがわかる .

基本的にスプライトモードは “Stefan” のようにカメラが同じ場所を何度も往復して撮像する場合に最も威力を発揮する . この場合 , シーケンスが長ければ長いほど , 1 秒あたりのビット数は少なくなる . また , “Horserace” や “Athlete” のように一方向にパンをしている画像は , スプライトの大きさはシーケンスの長さ按比例して大きくなる . よって , 特にほぼ定速でパンしている場



図 5.9: 背景スプライトの抽出例 . (a)Horserace , (b)Board , (c)Athlete

合などは1秒あたりのビット数はどの長さにシーケンスを区切ろうとも一定の値になる。しかし、この場合でも劇的な符号量削減が可能であることが実験結果が示している。

ところで、画像”Stefan”は背景の観客が無視できないほど動いているので、本来前景として抽出されるべき領域が前景比率の制限のため抽出されず情報が欠落し、その結果情報量が削減された可能性が高い。しかし、その他の”Stefan”を除いた画像は、視覚的にはカメラ操作による動き以外の変化はほとんど検知できない。しかし、完全なる自然画像であるので、1,2画素程度微妙に動いているものもあると考えられる。よってこれらの微小領域は制限前景比率内では抽出されず、符号化されていない。しかし、これら領域はごく微小であり、仮にこれらの領域を前景に回復したとしても前景比率には大方影響が出ないと考えられる。よって、劇的な符号量削減は情報の欠落がもたらしたものではないということが推測できる。

処理時間は、動きベクトル算出、画像処理（GM算出以降）、符号化全体の処理時間の合計を100%とすると、それぞれ78%、18%、4%であった。画像処理、符号化にかかる処理量は動きベクトル算出時間に比べて小さいことがわかる。

5.3.3 同一ビットレート符号化時のフレームレートと画質における評価

本節では実際に低レートでの符号化を想定し、上記の”Horserace”画像にて、ビットレート128kbps、初期フレームレート15fps、および、ビットレート64kbps、初期フレームレート10fps

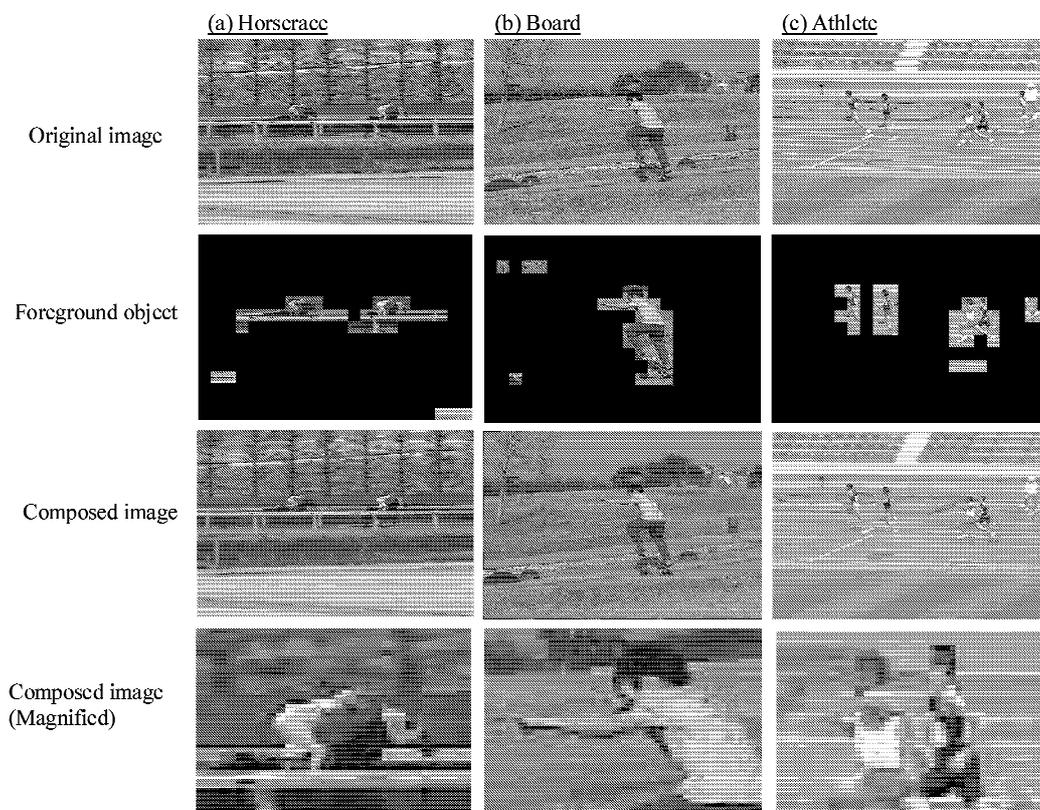


図 5.10: 原画像，前景オブジェクト，合成画像の例．(a)Horserace，(b)Board，(c)Athlete

にてノーマルモードとスプライトモードで符号化実験を行った．ノーマルモードの符号化制御およびスプライトモードの前景オブジェクトの符号化制御は VM Ver.15[13] に基づく．スプライトモードの前景比率は前節の知見より，15%とした．評価はフレームレートならびに SNR による客観画質による．

また，図 5.12 に 128kbps，64kbps における各画像のフレーム毎の SNR を比較した様子を示す．図 5.13 にノーマルモードとスプライトモードにおける符号化画像の例を示す．128kbps において，ノーマルモードは平均 SNR が 25.85[dB]，達成フレームレート 6.80fps に対し，スプライトモードでは，平均 SNR が 25.69dB，達成フレームレート 15fps であった．同様，64kbps においては，ノーマルモードの平均 SNR が 26.31[dB]，達成フレームレートが 2.8fps に対し，スプライトモードでは平均 SNR が 24.79[dB]，達成フレームレート 8.60fps であった．同一ビットレート下において，スプライトモードがノーマルモードの倍以上のフレームレートを達成できることが確認された．

さらに，前節の図 5.8 の”Hoserace”のスプライトモードの合成画像と本節の図 5.13 の”Horserace”の両モードの画像は明らかに前者の方が主観画質はよいが，SNR による客観画質は同程度であ

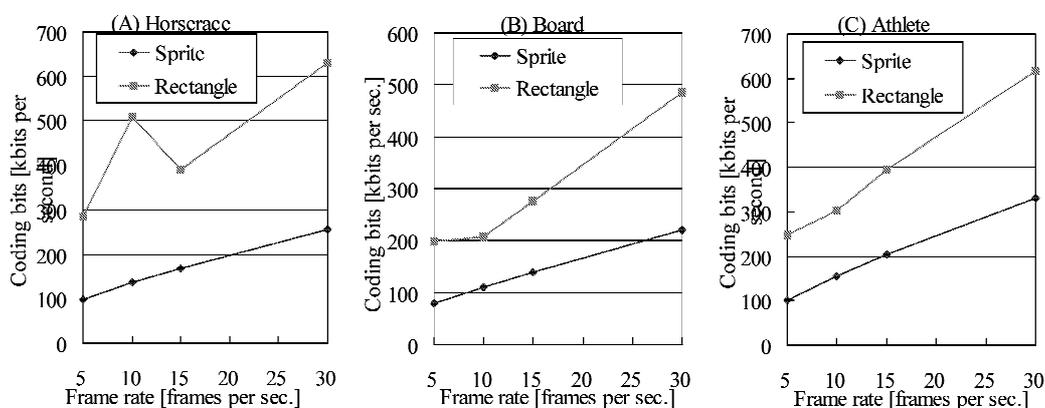


図 5.11: ノーマルモードおよびスプライトモードにおけるフレームレートと符号量の関係。

る．このことから、SNR から単純に主観画質を述べることはできないことを記しておく．また、前節の実験では、前景比率が 10-15 % 程度の条件で、スプライトモードが同程度の主観画質をノーマルモードの 1/2-1/4 の符号量で達成できると報告した．本節では、同条件の下、スプライトモードが同程度の符号量にて、同程度の主観画質、倍以上のフレームレートを達成したこと述べた．双方の実験結果はスプライトモードが前景比率がある程度小さいという条件の下、ノーマルモードに比較して劇的な圧縮効率を得るという点で一致する．

5.4 本章のまとめ

インターネットやモバイル通信環境下における超低ビットレートの符号化方式の開発を目的として、本論文では、スプライト符号化の課題の一つであるビデオオブジェクトの自動抽出アルゴリズムを提案した．また、提案アルゴリズムを MPEG-4 スプライト符号化に適用して MPEG-4 通常符号化方法と比較、検討を行った．前景比率が画像全体の 10-15 % 程度に収まる条件下で、MPEG-4 通常符号化方法の 1/4-1/2 以下の符号量で同程度の画質を実現した．また、実際に超低レート (128kbps, 64kbps) で符号化実験し、同程度の SNR においてスプライトモードがノーマルモードの倍以上のフレームレートを達成できることがわかった．

このように、スプライト符号化は特定の条件の下で、劇的な符号量削減が期待できる．しかし、すべてのビデオシーケンスがこの条件を満たしているわけではなく、スプライト符号化できるショットを判定する方法の検討が必要である．また、実際に配信を考慮した場合のビデオオブジェクト毎符号量割り当ての検討が今後の研究課題である．

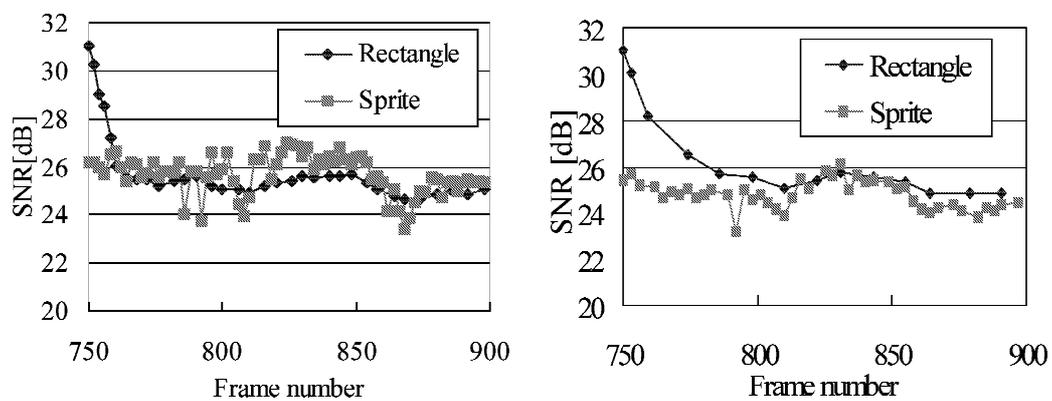


図 5.12: 同一ビットレート符号化時の SNR による画質評価 . 左 : 128kbps , 右 : 64kbps.



図 5.13: 128kbps におけるノーマルモードおよびスプライトモードの符号化画像の例 . 上段 : ノーマルモード , 下段 : スプライトモード .

第6章 スプライトを適応的に利用したMPEG-4 映像符号化システム

6.1 はじめに

筆者は、前章までに、MPEG-4 Main Profile[3]における「スプライト」という新しい符号化ツールに着目し、動画像におけるカメラ操作¹のみを反映した背景領域にスプライトを適用する方法を検討してきた。スプライトとは、パラメトリックな動きをする動画像の一連の領域を一枚の静止画像（パノラマ画像，モザイク画像）に射影したもので，元の領域はスプライトを幾何変換することで得られる [9][10][12]。筆者らは，スプライトが効果的に働く場合は，従来符号化法に比べて劇的な符号量の削減をもたらすことを明らかにしてきた [57]。同時に，スプライトがあらゆる動画像に効果があるとは限らないことも知られており，適応的な利用が望まれていた。スプライトの生成及び符号化の研究の歴史は長く，1990年代以降，Irani[9][10]やWang[12]は，フレーム間の大局的な動き（グローバルモーション）の抽出とスプライト（モザイク画像）生成について検討を行っている。また，Lee[35]やSmolic[15]らはスプライトを符号化に適用する研究を行っている。しかし，これらの研究はあらかじめビデオオブジェクトとして抽出された背景スプライトを符号化方式に適用するのに留まり，汎用的な符号化システムにスプライトを導入するための検討はなされていなかった。

本章では，動画像の圧縮率向上を目的とし，スプライトを適応的に利用した MPEG-4 Main Profile 準拠の汎用的な高圧縮映像符号化システムについて述べる。本符号化システムを構築するにあたって，以下の2点において新たな提案を行う。

一つは，スプライトが劇的な圧縮効果を生み出す場合もあれば，逆に符号化効率を悪化させる場合もあるという特性 [57][58] を鑑み，動画像からスプライトが効果を発揮する動画像の一連のフレーム群（便宜上，これをスプライトショット²と呼ぶ。）を自動的に抽出する方法を提案する。符号化ツールであるスプライトを適用するには，フレーム単位ではなく，ショット単位に適用性を判断する必要がある。符号化ツールの適応的選択の例は，古くは，マクロブロックをフレーム内予測とフレーム間予測を切り替えて符号化する方法や，ピクチャタイプ³を適応的に切り替える方法 [7] の提案がなされている。しかし，これらの方法は動き補償 (MC) を基本ツールとし，基本的にフレーム毎においても判定が可能である。一方で，スプライトショットの抽出に関しては，筆者ら [59] が先んじて提案してきたが，スプライトの効果のある要因の追求およびその組み合わ

¹パン，チルト，ズームといったカメラの動き。

²一連のフレーム群をショットと言う。

³I,P,B ピクチャ等。

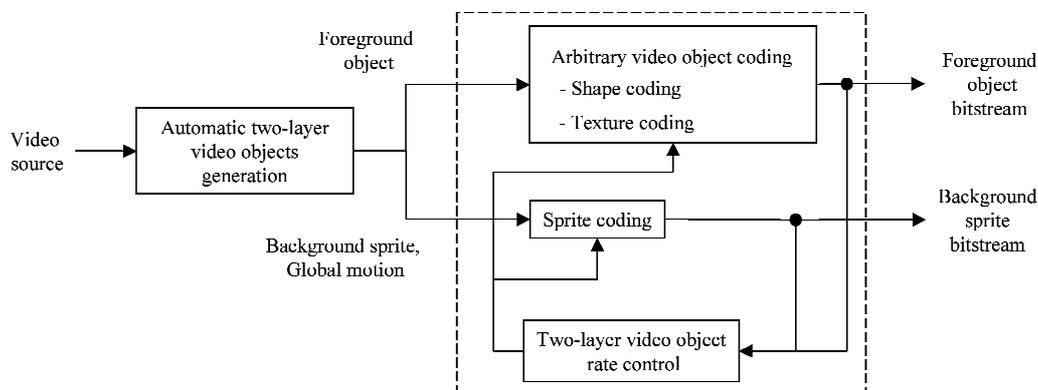


図 6.1: スプライトショットの概念 .

せによる検討に留まり、実際に符号化システムに導入するためのアルゴリズムの検討はなされていなかった。本章では、フレーム単位で適応性を判断する従来技術とは異なり、ショット単位にスプライトの適応性を判定し、汎用符号化システムへの導入を試みる。

二つ目は、2層ビデオオブジェクトレート制御法の提案である。MPEG-4 Main profile においては、複数のビデオオブジェクトを個別に符号化することが可能である。スプライトも一種のビデオオブジェクトである。複数の任意形状のビデオオブジェクトを同時にレート制御する方法として、Vetro[61] がマルチビデオオブジェクトレート制御法の提案を行っている。しかし、任意形状ビデオオブジェクトとは異なる性質を持つスプライトには適用できなかった。本章では、ストリーミング配信への適用も考慮し、低遅延スプライトと任意形状ビデオオブジェクトからなる2層ビデオオブジェクトのレート制御法を提案する。

以降、6.2 節では本章における符号化システムの概要を述べ、6.3 および 6.4 節では、本システムを構成するにあたってスプライトショット抽出法および2層ビデオオブジェクトレート制御法をそれぞれ新規に提案する。6.5 節では符号化実験による提案システムの評価を行い、6.6 節で結論を述べる。

6.2 スプライトを適応的に用いた符号化システムの提案

本章では、スプライトを適応的に用いた映像符号化システムの提案を行う。以下に、スプライトショット符号化の概念、スプライトが効果的な条件、および、本提案符号化システムの概略について述べる。

6.2.1 スプライトショット符号化

スプライトを使った符号化をスプライト符号化という。筆者らは文献 [57] にて、画像を前景オブジェクトと背景スプライトに自動的に分離し、背景スプライトにスプライト符号化を、前景オ

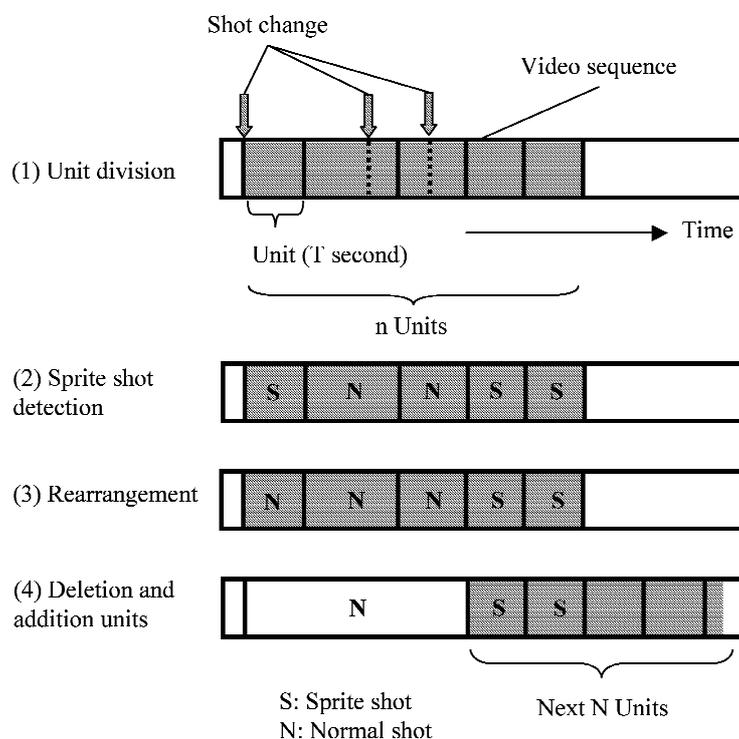


図 6.2: 本提案符号化システムにおけるユニット処理の概念。

プロジェクトに任意形状ビデオオブジェクト符号化を適用する，スプライトショット符号化⁴を提案している．図 6.1 にスプライトショット符号化の概念を示す．生成された 2 つのビデオオブジェクトのビットストリームは多重化されて送信される．デコーダ側では多重化されたビットストリームを多重分離し，各々のビットストリームを復号し，合成して表示する．一方，これに対して，スプライトを使用しない，MC+DCT を基本とする符号化をノーマルショット符号化と呼ぶことにする．

スプライト符号化のビットストリーム（符号化データ）は，一枚の背景スプライトを静止画符号化したデータと，フレーム毎のグローバルモーションデータからなる．また，オブジェクト符号化のビットストリームは，形状符号化データとテクスチャ符号化データからなる．スプライトは連続するフレームの一連の領域が一枚の静止画であらわされるので，その領域の占める割合が画面に対して大きいほど劇的な圧縮効果をもたらす．しかし，高い主観画質を従来符号化法に比べて小さいデータ量で表現できるが，元の領域のデータを忠実に再現するわけではないので，元データと復号データの PSNR は劣化するという特徴をもつ．

⁴文献 [57] ではスプライトモードと呼んでいる．

6.2.2 スプライトショット符号化に適したショットの条件

以下に、スプライトショット符号化が効果的に働く動画像の条件 [59][57][58] として、以下の項目を挙げる。

1. ショットチェンジを含まないこと。
2. グローバルモーションが十分大きいこと。
3. 前景領域が画面全体に対して十分小さい領域であること。

グローバルモーションとは、画像の大局的な動きである。グローバルモーションがほとんど無い場合は、前のフレームの同じ位置にあるマクロブロックの値をそのまま利用する”Not Coded”マクロブロックが多数出現する [59]。よって、従来方式による符号化でも十分な画質を得ることが可能である。また、前景オブジェクトの画面全体に対する割合が 30 %を超えると、スプライトショット符号化の効率が MC+DCT を基本とする従来符号化方式より劣化する [57]。

6.2.3 ショット単位での処理の流れ

スプライトの適用性を判定するにはショット単位での処理が必須である。よって、画像データは T 秒単位のフレーム群（これをユニットと呼ぶ）毎に処理される。本符号化システムは、ユニット毎にスプライトショットを逐次検出しながら、同時に符号化処理も行う。図 6.2 にユニット処理の概念を示す。

1. ユニット分割処理部において、 T 秒毎のユニットに分割する。次の t 秒 ($t < T$) 以内にショットチェンジがある場合は、ユニットの長さを $t + T$ 秒に延長する。前ユニットがショットチェンジで終わっていて、かつ、 t ($t < T$) 秒以内にショットチェンジがある場合は、ショットチェンジをまたぐ T 秒を強制的にノーマルショットとする。(図 6.2-(1))
2. ユニット数が n 個になったところで一旦画像の読み込み処理を停止する(図 6.2-(2))
3. n 個のユニットにおいて、スプライトショット抽出を行う。
4. 孤立するノーマルショットおよびスプライトショットを除去するために、判定結果に対し、ノーマルショットで囲まれた単独で存在するスプライトショットをノーマルショットへ変更する(図 6.2(3))
5. 1 個以上 $n - 1$ 個以下のユニットの判定を確定し、スタックをあける。空いたスタックにユニット数が n 個になるまで映像データを読み込む(図 6.2(4))
6. 以上を繰り返す。

このように、スタックには常に n 個のユニットが格納され、1 個以上 $n - 1$ 個以下のユニットから成るショットの判定が随時行われる。判定後のユニットはスタックから削除され、削除された数のユニットが新たに挿入される。

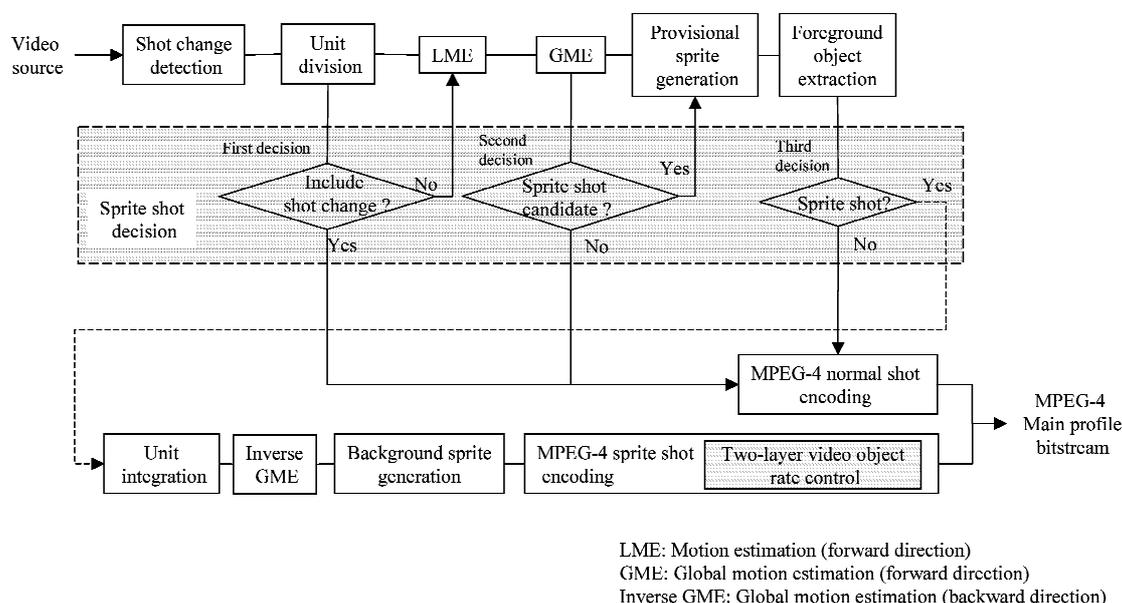


図 6.3: 本提案符号化システムのブロックダイアグラム。

6.2.4 提案符号化システムの概略

図 6.3 に、提案符号化システムのブロックダイアグラムを示す。スプライトショット抽出部と 2 層ビデオオブジェクトレート制御部が本符号化システムの主な新規な部分である。以下に提案符号化システムのブロックダイアグラムの説明を記す。

入力された画像に対して、まずショットチェンジを検出する。それと同時にユニット分割を行い、「5.2 スプライトショット符号化に適したショットの条件」で述べた 3 つの条件の一つである、(1) ショットチェンジの有無による 1 次判定を行う。

一次判定にてノーマルショットと判断されたユニットは、以降の判定処理を省略し、ノーマルショット符号化する。スプライトショット候補と判定されたユニットに対して、局所的な動きベクトルを算出し (LME)、それを基に順方向のグローバルモーションを算出する (GME)。

ここで、(2) グローバルモーションの有無による 2 次判定を行い、ノーマルショットと判定されたユニットは以降の判定処理を省略し、ノーマルショット符号化する。スプライトショット候補と判定されたユニットに関しては、仮スプライト算出し、それを利用して前景オブジェクトを算出する。

ここで、(3) 前景比率の大きさによる 3 次判定によってスプライトに適したユニットであることを判定する。前景比率とは、画面全体の面積に対する前景オブジェクトの面積比率である。ノーマルショットと判定されたユニットは、ノーマルショット符号化される。ここでスプライトショットとして判定されたユニットは、前述のショット内処理で示したように、2、3 の連続するスプライトショットはひとつのスプライトショットとして統合する (Unit integration)。また、孤立し

たスプライトショットはノーマルショットとして再判定され、ノーマルショット符号化をされる。複数のユニットを統合したスプライトショットは、スプライト生成に必要な逆方向グローバルモーションの算出 (Inverse GME) と背景スプライトの生成を行い、スプライトショット符号化を行う。

6.3 スプライトショット抽出アルゴリズム

本章では、本符号化システムの、スプライトショット抽出アルゴリズムの詳細を述べる。図 6.3 からスプライトショット抽出処理を列挙すると、以下の手順になる。

1. ショットチェンジの抽出
2. ショットチェンジの有無による 1 次判定
3. 順方向グローバルモーション算出
4. グローバルモーションの大きさによる 2 次判定
5. 仮スプライト生成
6. 前景オブジェクト抽出
7. 前景比率による 3 次判定

本提案符号化システムにおけるグローバルモーション算出、前景オブジェクト抽出法、仮スプライトおよび背景スプライト抽出は文献 [57][51] の方法を踏襲する。

以下に、グローバルモーションの大きさによる 2 次判定 (手順 4) および、前景比率による 3 次判定 (手順 7) の詳細を示す。

6.3.1 グローバルモーションによる 2 次判定

本章では、パン、チルトの組み合わせから 2 次判定を行う。ユニット内の先頭フレームを基準フレームとして、手順 3 により順方向にグローバルモーションを算出する。ここで、基準フレームの座標系を基本座標系とする。各フレームにおける各座標はグローバルモーションによって基準座標系の座標にマッピングすることができる。ここで、各フレームの中心座標 (0, 0) を基準座標系へのマッピングする。マッピング後の座標は、各フレームの x 軸および y 軸方向への変移、すなわち、パンとチルトの値に相当する。このパン、チルトの基準座標系における時間的な変移をトラジェクトリと呼ぶ。ここで、ユニット内のトラジェクトリの大きさ算出し、この大きさをを用いて 2 次判定を行う。判定方法の詳細を以下に述べる。

1. ショット内の各フレームの中心座標において、グローバルモーションによるトラジェクトリを算出する。
2. 基本座標系における、トラジェクトリの x 軸および y 軸方向の最大値、最小値 SX_{max} , SY_{max} , SX_{min} , SY_{min} をそれぞれ算出し、その差分絶対値 SX , SY を算出する。

$$SX = abs(SX_{max} - SX_{min}) \quad (6.1)$$

$$SY = abs(SY_{max} - SY_{min}) \quad (6.2)$$

3. $SX \geq Th_{tri}$ もしくは、 $SY \geq Th_{trj}$ を満たす場合はスプライトショット候補、満たさない場合はノーマルショットと判定する。ここで、ユニット内のフレーム数を $frames$ として、 $Th_{tri} = Th_{trj} = frames * 2.0$ とした。これは、フレームあたりのx方向およびy方向の動きが平均して2画素以下の場合をグローバルモーションが十分に小さいとして、ノーマルショットと判断することを意味する。

6.3.2 前景比率による3次モード判定

本小節では、前景比率による3次モード判定を提案する。前景オブジェクトの抽出は文献[57][58]を踏襲する。文献[57]にて、画面に対する前景の面積（前景比率）が10-15%程度のところでスプライトが効果的に作用し、30%を超えると、逆にスプライトの効率が通常の符号化に比べて悪くなることを示されている。そこで、前景比率によって、スプライトが適用できるか否かを判定する3次判定を行う。ユニット内の前景比率の平均値が十分小さくても、ショット内には前景の面積が急激に増えている場合がある。これは、例えば、カメラの前を突然人が横切るなどの場合に起きる。この現象を回避するために、前景比率が極端に大きくなっているフレーム数を算出し、これが全フレーム数のある割合以上を占めるショットをノーマルショットと判定する。以下に、3次判定の流れを記す。

最初に、ユニット毎に前景オブジェクトおよび前景比率を算出する。前景の大きさを fg_{size} （単位：画素）、フレームの大きさを $frame_{size}$ （単位：画素）とすると、前景比率 FG_{ratio} は以下の式で表される。

$$FG_{ratio} = fg_{size} / frame_{size} \quad (6.3)$$

さらに、ユニット内の平均前景率を $AVEfg_{ratio}$ 、ユニット内のフレーム数を $frames$ 、スプライトショットとして許容できる前景比率の最大値の閾値を Th_{max} 、前景比率が極端に大きくなっているフレームの割合の閾値を Th_{large} として、以下のように判定する。

1. ユニット内の平均前景率 $AVEfg_{ratio}$ を算出。
2. ユニット内の各々のフレームの前景比率 FG_{ratio} が Th_{max} を超えるフレーム数 (count) を算出する。
3. 以下の判定法で判定する。“S”はスプライトショット，“N”はノーマルショットである。
 - if $AVEfg_{ratio} \geq Th_{ratio}$ mode=N;
 - else if $(count \geq frames * Th_{large})$ mode=N;
 - else mode=S;

表 6.1: 実験画像

Image	Frames	size	contents
horserace(a)	999	SIF	pan, zoom
horserace(b)	999	SIF	pan, zoom
skateboard	390	SIF	pan, tilt
stefan	300	SIF	pan, zoom

後述の評価実験では, $Th_{max} = 0.2(20\%)$, $Th_{large} = 0.2(20\%)$, $Th_{ratio} = 0.17(17\%)$ を用いた.

6.4 2層ビデオオブジェクトレート制御法

スプライトショットは, 前景オブジェクトと背景スプライトの2層のビデオオブジェクトを符号化する. よって, Verification Model Ver. 17.0[62] 等のレート制御法をそのまま適用することはできない. また, Vetro[61] らが提案したマルチビデオオブジェクトレート制御法はスプライトには適用できない. また, スプライトをストリーミング配信するためには, 低遅延スプライト符号化を行う必要がある. よって, 本章では, 低遅延スプライトに対応した2層ビデオオブジェクトレート制御法 [60] を提案する.

前景オブジェクトと背景オブジェクトの符号量比率に基づいて符号量を割り当て, 低遅延スプライトを用いて符号化を行う. 具体的には以下の手順で処理を行う.

1. スプライトショットにおいて, 量子化値 $QP = 15$ (QP : Quantization Parameter) にて前景オブジェクトおよび背景スプライトを符号化し, その配分を前景背景比とする.
2. ショット全体のトータルビット $T[\text{bit}]$ と前景背景比から, 背景スプライトに割り当てられる符号量 $T_{b0}[\text{bit}]$ を算出する.
3. 背景スプライトの QP 値を 1 から 31 まで変化させて QP 固定 (1 フレームにひとつの QP 値) で符号化し, 前項で求めた背景スプライトに配分される符号量 $T_{b0}[\text{bit}]$ に最も近くなる QP 値 QP_b およびその符号量 $T_b[\text{bit}]$ を求める.
4. 残りのビット数を計算し, 前景オブジェクトの総ビット数 $T_f = T - T_b[\text{bit}]$ とする.
5. $T_f[\text{bit}]$ を秒数で除した値を目標ビットレートとして, 前景オブジェクトを VM17.0[?] で符号化する. これにより, 前景のフレームレート $f_f[\text{fps}]$ が算出される.
6. スプライトを前景のフレームレート $f_f[\text{fps}]$ に合わせて, $QP = QP_b$ (固定) にて低遅延モードでスプライト符号化する.

表 6.2: スプライトショット抽出結果の一例 (”skateboard”)

Start frame	End frame	Second decision	Third decision	FG ratio	Final decision	Final shot
1	30	N	N		N	N
31	60	N	N		N	
61	90	N	N		N	
91	120	S	N	0.214	N	
121	150	S	S	0.092	S	S
151	180	S	S	0.101	S	
181	210	S	S	0.120	S	
211	240	S	S	0.063	S	S
241	270	S	S	0.097	S	
271	300	S	S	0.106	S	S
301	330	S	S	0.093	S	
331	360	N	N		N	N
361	390	N	N		N	

6.4.1 符号化実験

前述の5つの画像で符号化実験を行った。比較対象は以下の2つである。

- MC+DCT を基本とする従来符号化法 (MPEG-4 Main profile 準拠 VM17.0[62])
- スプライトを用いた本提案符号化法 (MPEG-4 Main profile 準拠)

6.5 提案符号化システムの評価

提案符号化システムを表 6.1 の画像を用いて、PSNR、達成フレームレート、主観画像によって評価した。

6.5.1 評価画像

画像はカメラ操作を含む 10-30 秒程度のスポーツ映像 (競馬, スケートボード, サッカー, テニス) である。以下に画像の具体的内容を示す。

表 6.3: 平均前景率, スプライトショット符号化選択率.

Image	Sprite mode ratio	Foreground ratio
horserace(a)	0.478	0.140
horserace(b)	0.775	0.083
skateboard	0.538	0.096
stefan	1.000	0.094

horserace(a),(b)

”horserace(a)”は競馬のスタートから手前の直線コース, 第一コーナーを回るまで, ”horserace(b)”はその続きの第二コーナーから奥の直線コース, 第三コーナーに至るの競走馬をフォローショットしたものである.

skateboard

カメラがほぼ静止した状態で, 画面右端からスケートボードが登場, 画面の左に向けて移動する様子をカメラが追従したものである. 最後にスケートボードが向き逆方向に変えて画面から消える. カメラはほぼ静止する.

stefan

標準画像である. 左右にボールを拾うテニスプレーヤをフォローショットし, 最後は画面の右端から左にかけて移動する様子をカメラが高速に追従している.

本提案符号化システムにおけるノーマルショットはBピクチャを含まない. ユニット分割処理におけるショットの区切りは $T = 1$ 秒, またユニット数 $n = 5$ とした. また, 目標ビットレートおよび目標フレームレートを以下に示す.

- 128kbps, 15fps
- 384kbps, 30fps

表 6.2 にスプライトショット抽出結果の例(”skateboard”)を示す. ”second decision”はグローバルモーションの大きさによる判定, ”Third decision”は前景比率による判定である. フォローショットの部分がスプライトショットとして抽出されている. また, 表 6.3 にスプライトショット符号化の選択率と平均前景率を示す. ”stefan”は標準画像として知られるが, その平均前景比率は約 10%, すべてのショットにてスプライトショット符号化が選択されている. ”horserace(a)”と”horserace(b)”に着目すると, ”horserace(a)”の方が, 静止している時間が長く(スタート時),

表 6.4: 符号化結果

Image	Method	Object type	128kbps, 15fps			384kbps, 30fps		
			Ave. QP	Ave. SNR [dB]	Frame rate [fps]	Ave. QP	Ave. SNR [dB]	Frame rate [fps]
Horse(a)	Proposed	FG	19.99	25.35	15.06	10.51	29.45	30.00
		BG	20.42	19.76	15.06	10.31	19.82	30.00
		Norm.	30.10	24.80	10.88	23.77	25.90	30.00
		Total	—	23.24	12.88	—	24.21	30.00
	Conv.	—	30.37	24.97	8.83	25.12	20.20	30.00
Horse(b)	Proposed	FG	14.41	28.06	15.00	6.91	33.43	29.65
		BG	14.27	22.74	15.00	7.07	22.99	29.65
		Norm.	17.36	29.31	14.53	10.90	32.32	30.00
		Total	—	25.39	14.89	—	26.50	29.73
	Conv.	—	27.00	26.94	7.39	14.12	29.40	30.00
Skate	Proposed	FG	18.97	24.29	15.00	9.36	30.07	27.57
		BG	18.71	20.29	15.00	9.62	20.25	27.57
		Norm.	24.07	14.46	14.83	10.95	26.83	30.00
		Total	—	23.16	14.92	—	24.46	28.69
	Conv.	—	26.40	24.46	12.85	18.15	26.38	30.00
Stefan	Proposed	FG	17.07	26.44	15.00	8.34	31.48	30.00
		BG	17.00	18.19	15.00	8.30	18.23	30.00
		Norm.	—	—	—	—	—	—
		Total	—	19.38	15.00	—	19.52	30.00
	Conv.	—	30.84	24.15	8.90	23.67	25.37	3 0.00

またカメラに近いところを競走馬が走ることから，前景比率が高く，”horserace(b)”に比較しスプライトショットになる比率が低い．また，スプライトショットにおける平均前景比率も高い結果となっている．以上，表 6.2, 6.3 から，本スプライトショット抽出が効果的に作用していることが分かる．また，図 6.4 に”skateboard”におけるフレームにおける平均 QP 値の推移を示す．表 6.2 から，従来符号化法にて QP 値が 30 前後に張り付いている場合でも，提案手法では，スプライトショット符号化が採用され，前景・背景とも 10～15 程度の QP 値が与えられており，より細かい量子化が実現されている．すなわち，従来手法では，スケートボードをフォローショットする部分では十分な符号量が与えられず，量子化値が大きくなったのに対し，提案手法ではこの部分にスプライトショット符号化が用いられることによって，より高い画質（細かな量子化）を実現した．また，従来方式では，ショットの最初，中盤，後半では，あきらかに中盤に対して粗い量子化が行われていたのに対し，本提案手法では，安定した画質（量子化値）を実現していることがわかる．

図 6.5 にスプライトショットの合成画像の例を示す．表 6.4 に符号化効率，平均 QP (Quantization Parameter: 量子化パラメータ)，平均 PSNR (信号対雑音比)，達成フレームレートの比較の例を示す．”Mode”における”Proposed”は提案符号化法，”Conventional”は従来符号化法である．表 6.4 において，128kbps, 15fps の符号化条件の達成フレームレートに着目すると，”horserace(a)”, ”horserace(b)”, ”skateboard”, ”stefan”では，従来符号化では各々 8.33fps, 7.39fps, 12.85fps, 8.00fps

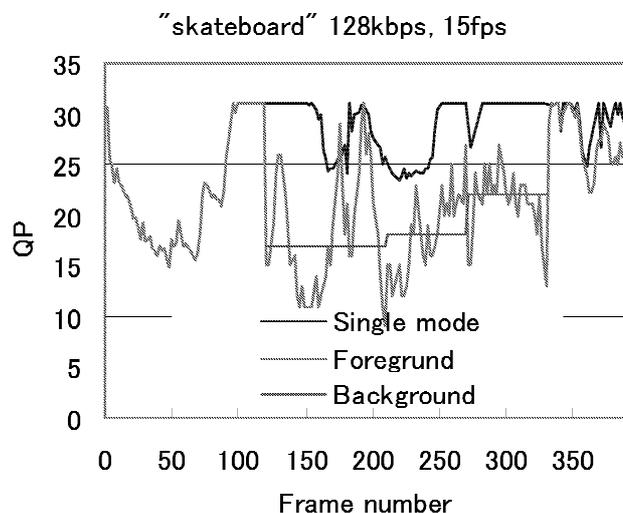


図 6.4: フレームにおける平均 QP の変化 .

のフレームレートであるのに対し，提案符号化法では，12.88fps 14.89fps, 14.92fps, 15.00fps の高フレームレートが達成されている．この符号化条件では，平均 QP 値に関しても，提案符号化法が従来符号化法より小さい値が割り当てられている．”horserace(b)” , ”skateboard” の 2 つの画像では，提案符号化法におけるノーマルショットのフレームレートと従来符号化法のフレームレートはほぼ同程度であることから，従来符号化法がフレーム落ちを起こす部分が提案符号化法ではスプライトショット符号化され，目標フレームレートに近いフレームレートを実現していることが分かる．

384kbps, 30fps の符号化条件では，従来符号化法と提案符号化法の両者ともおよそ 30fps のフレームレートを達成している．しかし，平均 QP 値に着目すると，圧倒的に，提案符号化法が小さい値が割り当てられており，より細かい量子化が行われている．”horserace(b)” , ”skateboard” の 2 つの画像では，提案符号化法におけるノーマルショットの平均 QP と従来符号化法の平均 QP の値がほぼ同じであることから，従来符号化法が十分な品質を与えられない部分が提案符号化法ではスプライトショット符号化され，平均的に高い品質を実現していることが分かる．

一方，提案符号化法におけるトータルの PSNR は従来符号化法の PSNR よりも低い値を示している．これは，背景スプライトが各々のフレームを貼り合わせて生成されたものであるため，そこから切り出す背景画像は元の信号を忠実に再現していないというスプライト符号化特有の現象である（本章第 1 節を参照）．しかし，前景オブジェクト（FG）に着目すれば，トータルの PSNR よりも 1 ~ 2 dB 高い PSNR を達成している．前景オブジェクトに割り当てられる QP 値が従来符号化法よりも小さい値が割り与えられおり，より細かい量子化が行われている．図 6.5 より，同じビットレートにて，提案符号化法の方が従来符号化法よりも高い主観画質を実現していることがわかる．



図 6.5: 符号化画像の例．左，提案方式 (本提案符号化システム)，右，従来方式 (VM 17.0)

以上のように，提案符号化法は，従来符号化法が不得意とするフォローショット画像において，高いフレームレートを達成し，細かい量子化による主観品質の向上をもたらす．よって，安定した品質を提供することが可能である．

図 6.6 に専用デコーダによってデコード表示されている様子を示す．また，図 6.7 にスプライトビューワによる原画像，本提案符号化システムによる符号化画像，前景画像，背景画像，グローバルモーションを表示した様子を示す．

本提案符号化システムは 1.5GHz Pentium IV プロセッサを有する汎用的な PC にて，実時間比にて 10 倍程度の処理時間でエンコードを行うことが可能であった．また，400MHz Pentium III 以上，256Mbyte 以上のメモリを有する汎用的な PC にて実時間デコードが可能である．

6.6 本章のまとめ

筆者は，スプライト符号化を用いた汎用的な MPEG-4 動画画像符号化システムを開発した．スプライトが効果的なショットを自動的に抽出する，スプライトイショット抽出アルゴリズムを提案した．また，前景オブジェクトと背景スプライトに自動的に符号量を配分するための 2 層ビデオオブジェクトレート制御法を提案した．また，実際に符号化実験を行って，低レートにおける従来符号化方式に対する本提案符号化システムの優位性を示した．

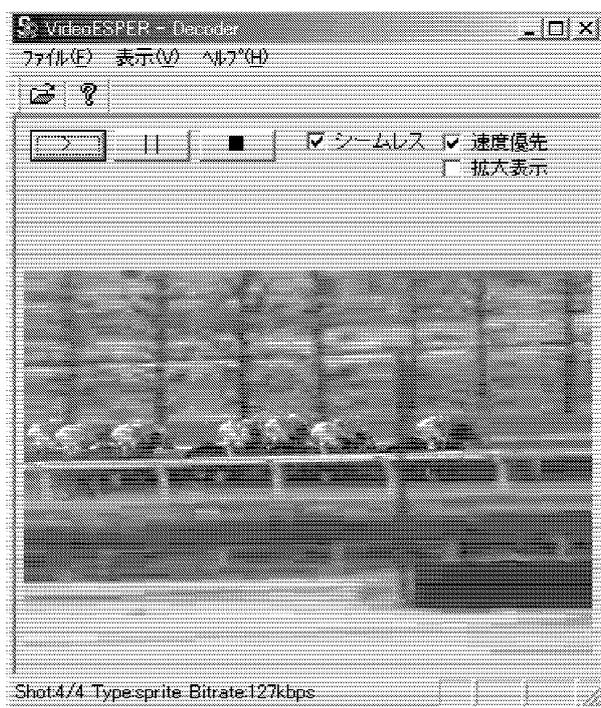


図 6.6: 提案符号化システム専用デコーダ.

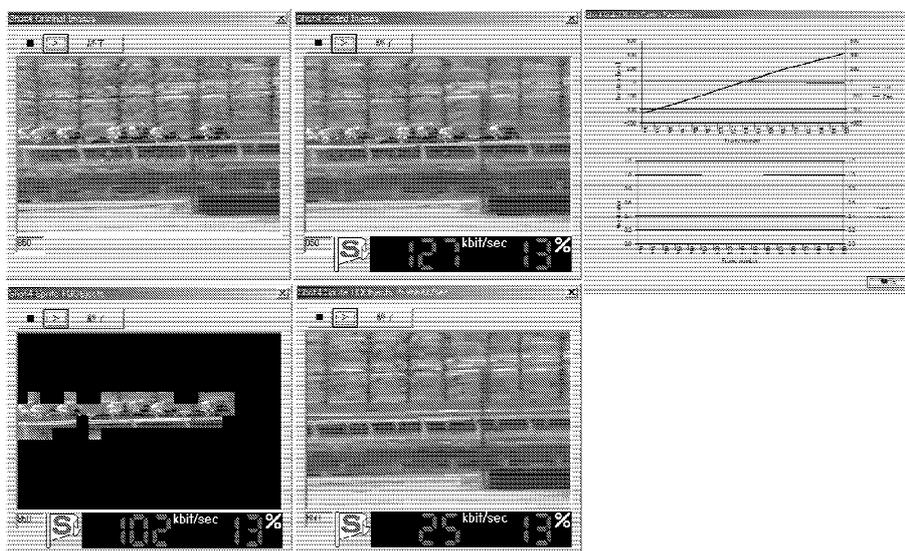


図 6.7: スプライトビューワによる生成オブジェクトおよびグローバルモーションの例．上段左より，原画像，符号化画像，グローバルモーション．下段左より，前景オブジェクト，スプライトから切り出された背景画像．

第7章 DCT係数を用いた主観品質に寄与する特徴の推定とスケーラブル符号化への適用

7.1 はじめに

インターネットやモバイル環境での符号化方式として、MEPG-4 Simple Profile (SP) ならびに Advanced Simple Profile(ASP)[3] が普及しつつある。これらの符号化方式は1Mbps以下のビットレートにてMPEG-1等の従来方式に比較して大きな圧縮効率を達成している。しかし、インターネットやモバイル通信環境の特徴であるパケットロスや帯域変動が起きた場合に適した符号化ではない。

一方、MPEG-4では、Fine Granularity Scalable (FGS) Profile[4]を拡張レイヤとしてスケーラブル符号化を実現するStreaming Profileを規定している。スケーラブル符号化とは、一つのビットストリームから複数のビットレートのビットストリームを取り出すことができる符号化方式である。このような利用をOne Source Multi Useという。スケーラブルなビットストリームは、伝送帯域の変動があった場合に、拡張レイヤの一部をカットして、画像が途切れなくすることが可能である。本章では、ビットストリームの一部がカットされても高い主観品質を維持するスケーラブル符号化方式の構築を目的とする。

拡張レイヤで用いられるFGS符号化は、原信号と基本レイヤのローカルデコード画像の差分信号をDCT変換した係数(差分DCT係数)をビットプレーン展開して、差分の大きいレイヤから順に符号化、伝送される。一般に、差分の大小は必ずしも主観的な画品質と関係があるとは限らない。

FGS符号化を用いたスケーラブル符号化における画質を向上させる方法として、Selective Enhancement(SE)および、Frequency Weighting(FW)という符号化優先順位を変更する手段がサポートされている。前者は優先的に符号化するマクロブロック(Macro Block:MB)を選択することが可能である。また、後者は全てのMBにおいて、DCT係数の低周波を優先的に符号化する。これらの適用方法は、標準化の対象外である。また、主観品質の良し悪しはその適用方法によるところが大きい。特にSEは(1)どの領域を、(2)どの程度優先させるかという2つの自由度を持ち、これらを適切に設定する必要がある。中嶋[45]らは、原画像の画素信号を用いて、SEを適用するMBを決定する方法の提案ならびにFGS符号化への適用を試みている。しかし、原画像を用いて判定するために、基本レイヤの画品質が反映された領域の抽出が行われていなかった。すなわち、SEの前者の自由度において、基本レイヤの品質に拠らず、常に同じ優先順位付けがなされる問題点があった。また、SEにおける後者の自由度について、優先したい領域をどの程度ビットプレーンをあげるべきかという検討が十分にされていなかった。

本章は、FGS 符号化において、画像の主観品質を左右する 2 つの歪であるブロックノイズおよびモスキートノイズが出現するブロックを差分 DCT 係数およびローカルデコード画像から推定する画像特徴抽出法を提案する。また、提案手法によって領域分割された結果を用いて、ビットプレーンをシフトするアルゴリズムを提案する。また、自然画像およびセルアニメーション画像にて、主観品質に寄与する領域を優先的に FGS 符号化することで、主観品質を向上させる条件について考察する。

以降、7.2 節にて FGS を用いた MPEG-4 スケーラブル符号化の概略、7.3 節にて差分 DCT 係数の特徴について、7.4 節にて主観画質に寄与する画像特徴の定義とその画像特徴を抽出するアルゴリズムの提案について、7.5 節では画像特徴抽出法によって領域分割された結果を用いたビットプレーンシフト方法の提案と MPEG-4 スケーラブル符号化に提案手法を適用した符号化実験について、7.6 節では、結論を述べる。

7.2 MPEG-4 スケーラブル符号化

7.2.1 MPEG-4 スケーラブル符号化器構成

図 7.1-(a) に通常の FGS 符号化器を示す。基本レイヤには Simple Profile もしくは Advanced Simple Profile が用いられ、拡張レイヤには FGS Profile が用いられる。基本レイヤにて符号化された画像と原画像を用いて差分画像を作成し、この差分値をブロック毎に DCT 変換する。この DCT 係数を差分 DCT 係数と呼ぶ。全ての MB における全ての差分 DCT 係数が 2 進数で表される。これをビットプレーン展開という。基本レイヤ、上位のビットプレーンから順次可変長符号化、伝送され、フレーム全体の画質が徐々に向上する。

図 7.1-(b) および (c) に本節で使用した符号化器の構成を示す。図 7.1-(b) の方法は Li[63] らがエンコーダの構成法の一例として述べている。FGS 符号化は、通常は図 7.1-(a) のように、原画像と基本レイヤにおけるローカルデコード画像の差分を符号化する。しかし、基本レイヤにおける量子化前の DCT 係数を DCT 逆変換するとほぼ原画像と同じ信号が再現されると仮定すると、図 7.1-(b) のように、量子化前の DCT 係数と量子化後の DCT 係数の差分（差分 DCT 係数）をそのままビットプレーン展開し、VLC 符号化することで、図 7.1-(a) とほぼ同じビットストリームを生成できる。図 7.1-(a) に比べて、ローカルデコード画像と原画像の差分を計算する処理と、その差分を DCT 変換する処理の 2 つの処理を省略できるため、演算量が少ないという利点がある。図 7.1-(c) は、図 7.1-(a) において、基本レイヤを縮小画像にて符号化し、そのローカルデコード画像を拡大したものと、原画像の差分を拡張レイヤにて符号化したものである。再生側では、基本レイヤのビットストリームを復号し、拡大処理をし、拡張レイヤのデコード画像と重畳する。この方法は、基本レイヤ縮小することで、画像によっては十分な符号量を与えることが可能である。基本レイヤの画質が（例えその一部でも）向上すれば、拡張レイヤに割り当てられる符号量が少ない場合でも、全体の画質が向上する可能性がある。本節では、図 7.1-(b) および (c) の構成にてスケーラブル符号化を行う。

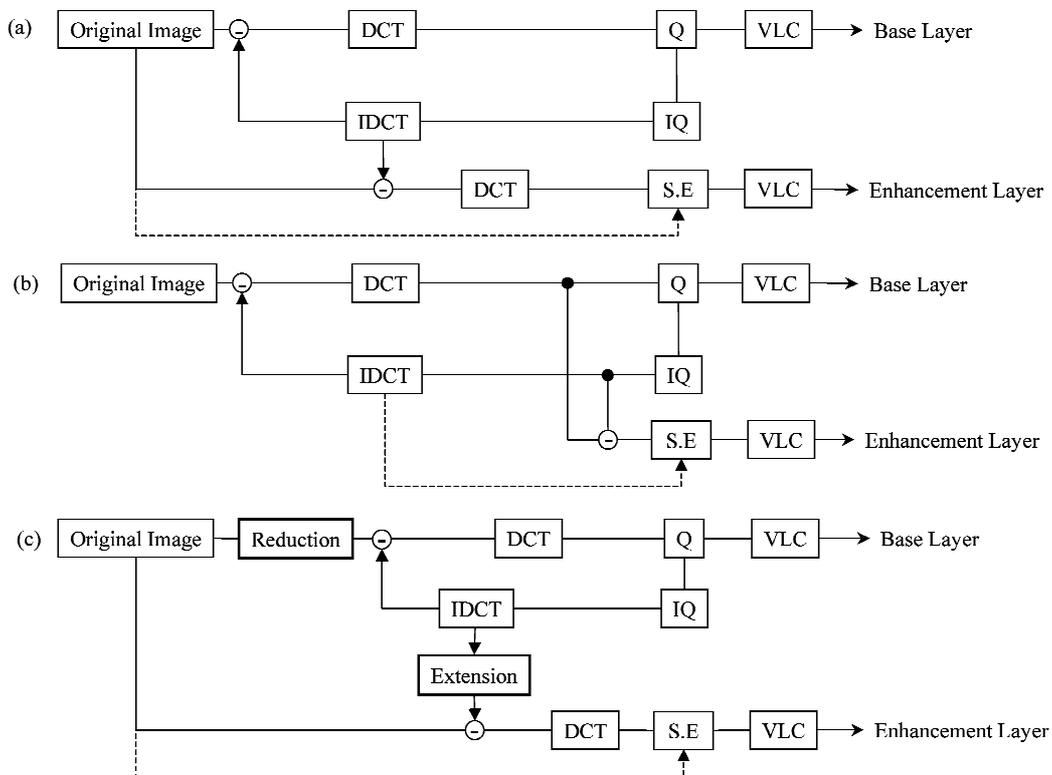


図 7.1: FGS を用いた MPEG-4 スケーラブル符号化器構成 : (a) 一般的な構成 (その 1) , (b) 一般的な構成 (その 2) , (c) 基本レイヤに拡大・縮小処理を加えた構成 .

7.2.2 FGS における Selective Enhancement

図 7.2 に SE の概念を示す . SE によって , MB 毎にビットプレーンのシフトを行うことが可能である . 差分 DCT 係数は図 7.2 のようにビットプレーン展開され , 上位プレーンから順番に符号化される . ここで , 任意に優先したい MB のビットを何ビットかシフトする . これによって , 差分 DCT 係数の振幅の小さな MB でも , SE によって上位のプレーンにて符号化 , 伝送される . どの MB をどの程度シフトするかは標準では規定されていない .

7.3 差分 DCT 係数と画像特徴

本節では , 差分 DCT 係数の分布と画像特徴の関係を考察する . レートコントロール下において , 参照画像が劣化したものを用いるか , ある程度高い品質のものを用いるかによって , 差分 DCT 係数の分布は異なってくる . すなわち , 差分 DCT 係数は基本レイヤの量子化ステップサイズに相当するパラメータである QP (Quantization Parameter) 値に影響される .

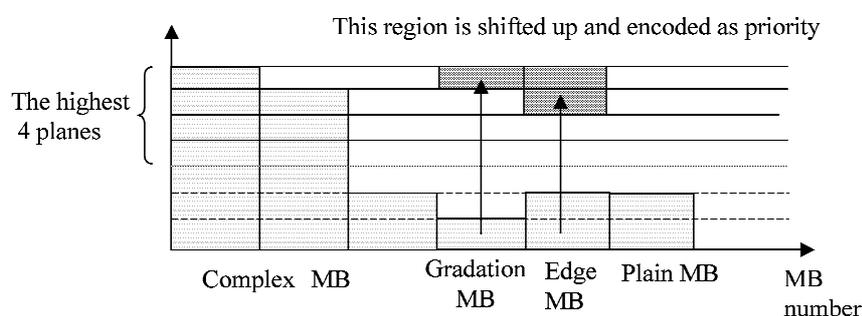


図 7.2: MPEG-4 FGS 符号化における Selective Enhancement の概念 .

7.3.1 差分 DCT 係数分布の特徴

差分 DCT 係数の分布は、ピクチャタイプによらず、基本レイヤの画質（量子化パラメータ値：QP 値）によって波形が異なる。

図 7.3 に画像”Flower garden”における P ピクチャの差分 DCT 係数分布の例を示す。横軸に低域成分から高域成分にジグザグスキャンを行った場合の DCT 係数番号，縦軸に係数の絶対値を記す。図 7.3-(a) は”花畑”のブロックの差分 DCT 係数，図 7.3-(b) は”空と屋根の境界”のブロックの差分 DCT 係数，図 7.3-(c) は”空”のブロックの差分 DCT 係数である。画像全体から見て，複雑なテクスチャの中にあるブロックを「複雑なブロック」，複雑なテクスチャと平坦なテクスチャの境界にあるブロックを「エッジを含むブロック」，平坦なテクスチャの中にあるブロックを「平坦なブロック」と定義すると，(a) は複雑なブロック，(b) はエッジを含むブロック，(c) は平坦なブロックと言い換えることができる。ここで，主観品質に影響を与えるモスキートノイズやブロックノイズが目立つのは (b) および (c) のブロックである。

平坦なブロックでは，QP 値によらず，ごく小さい振幅の差分 DCT 係数がどの周波数にも現れるという特徴を持つ。一方で，複雑なブロックとエッジを含むブロックについては，QP 値が大きいほど，振幅の大きな波形が現れ，逆に，QP 値が小さくなるほど，振幅は小さくなる。両者は同様の傾向を持つため，複雑なブロック，エッジを含むブロックの区別は困難である。一方で，平坦なブロックとそれ以外のブロックの区別は容易である。

7.3.2 平坦領域とそれ以外の領域の区別

ここで，量子化前の DCT 係数がある整数 N で除した時の商を計算し，商が”0”になった係数の個数をカウントする操作を行う。この操作は，閾値 N よりも小さい DCT 係数をカウントしたことに相当する。この数が $64(=DCT \text{ 係数の数})$ に近ければ近いほど平坦なブロックである可能性が高いと仮説を立てる。

以下の 4 つの画像に $N = 8$ の場合の操作を行い，その結果を図 7.4 に示す。

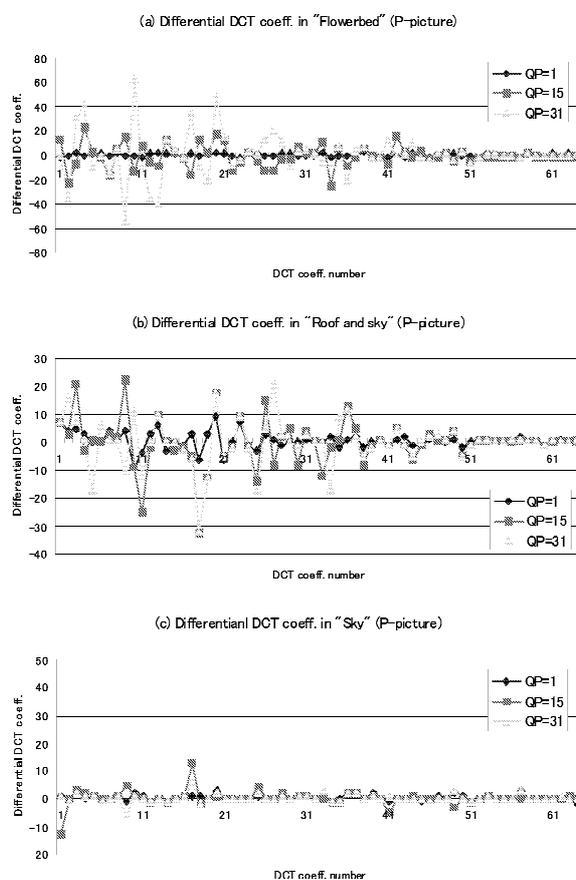


図 7.3: 基本レイヤにおける差分 DCT 係数の分布 (P ピクチャ) .

- ”Flower Garden” VGA(720 × 480 画素)
- ”Baseball” VGA(640 × 480 画素)
- ”Robot1” VGA(720 × 480 画素)
- ”Robot2” VGA(720 × 480 画素)

”Flower garden”および”Baseball”は自然画像, ”Robot1”および”Robot2”はセルアニメーション画像である. 図 7.4 において, 輝度値が高いブロックほど, 閾値 N よりも小さい DCT 係数の個数が多い. 図 7.4 より, ”Flower garden”の空の部分, ”Baseball”の背景部分, ”Robot1”および”Robot2”の均等色¹部分に高い値が, ”Flower garden”の花畑の部分, ”Baseball”の人物の部

¹均一な色で描かれた領域.

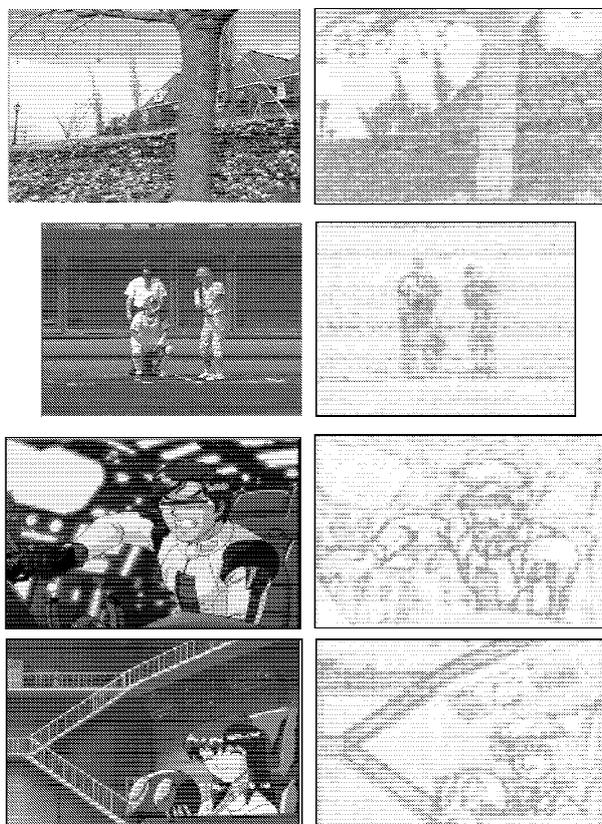


図 7.4: 差分 DCT 係数による領域分割例．上段より，“Flower Garden”，“Baseball”，“Robot1”，“Robot2”．

分，“Robot1”，“Robot2”の線画²に低い値が出現しており，仮説が正しいことを示す．以降，平坦なブロックを「平坦ブロック」，それ以外のブロックを「複雑ブロック」と呼ぶことにする．

7.4 主観画質に寄与する画像特徴抽出アルゴリズム

前節にて，平坦ブロックと複雑ブロックが差分 DCT 係数の分布特徴によって分類可能であることを示した．本節では，この2つのブロックの組み合わせを用いて主観画質に寄与する画像特徴を抽出するアルゴリズムを提案する．

7.4.1 主観画質を考慮した画像の領域の分類

主観品質への影響を基準にして，画像を以下の領域に分類する．

²均等色領域に対して，キャラクタの境界等を強調するために描かれた線．

- 画像の劣化が主観画質に大きく寄与する領域
 - － エッジ領域
 - － グラデーション領域
- 画像の劣化が主観画質にほとんど寄与しない領域
 - － 複雑領域
 - － 平坦領域

以下に、4つの領域の説明を記す。まず、エッジ領域は、十分な符号量が与えられない場合に、エッジの周辺にモスキートノイズが生じる領域である。主観画質に大きく影響を与えるブロックである。自身が複雑ブロックで周囲に平坦ブロックがある場合にそのブロックがエッジを含むブロックである可能性が高い。

グラデーション領域は、輝度がなだらかに変化する領域であるが、十分な符号量がない場合にブロックノイズが生じる領域である。主観画質に大きく影響を与えるブロックである。自身が平坦ブロックであり、かつ、周囲にも平坦ブロックがありその平均輝度値の差分が大きい場合がこれに相当する。

複雑領域は、自身も周囲も複雑な領域であるために、ノイズが目立ちにくいとされる領域である。一方で、この領域に属するブロックを優先的に符号化するとSNRの向上に大きく寄与するという特徴を持つ。

上記の3種類以外のブロックを平坦領域に属すると判断する。自身が平坦であり、周囲にも平坦なブロックがある場合、かつ、その輝度値にあまり変化がない場合に平坦領域に属する。少ない符号量でも高い品質を実現できる領域である。

7.4.2 画像特徴抽出アルゴリズム

図7.5に主観画質を考慮した画像特徴の抽出アルゴリズムを示す。本アルゴリズムは、差分DCT係数を用いて、画像を前節で定義した特徴を反映する領域に分類する。ただし、グラデーション領域については、そのブロックの平均輝度によって、高輝度グラデーション領域と低輝度グラデーション領域に更に分類する。以下に、主観品質に影響があると思われる順に、領域を列挙する。

1. エッジ領域
2. 高輝度グラデーション領域
3. 複雑領域
4. 低輝度グラデーション領域
5. 平坦領域

エッジ領域、高輝度グラデーション領域はそのノイズの出現が主観品質に大きな影響を与える領域である。複雑領域においては、フレームの一部に集中して現れる場合は、被写体等である可能

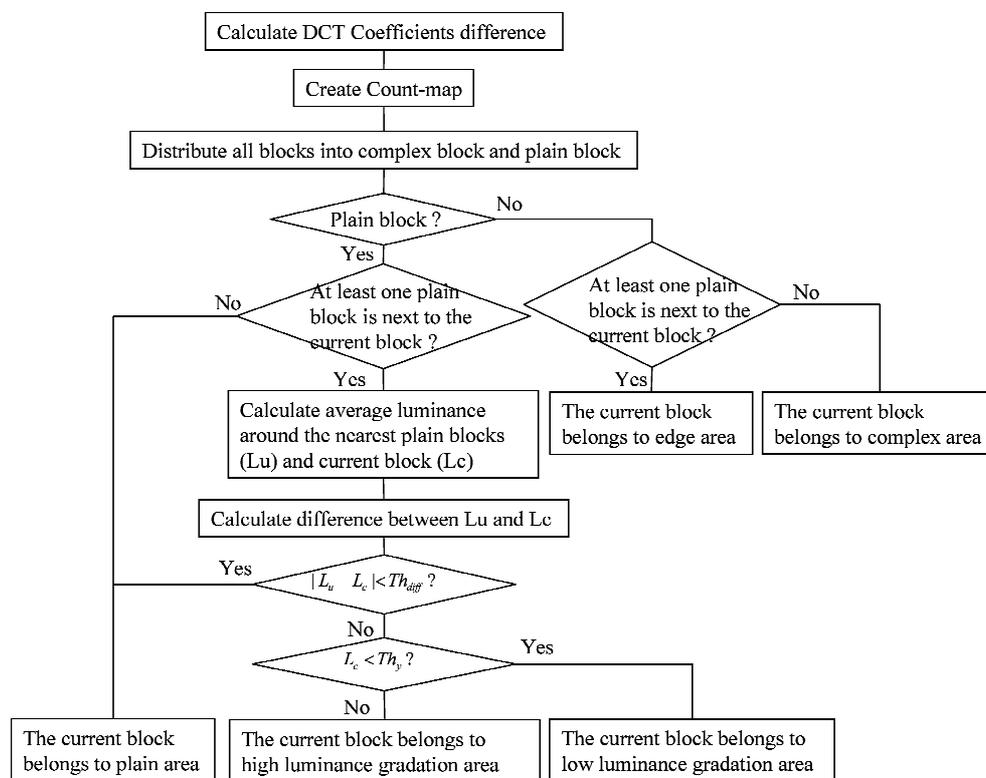


図 7.5: 差分 DCT 係数を用いた画像特徴抽出アルゴリズム .

性があり、この領域の品質の劣化は、主観品質に影響を与えられ、低輝度グラデーション、平坦領域ではノイズが主観品質に影響を与えにくいものとする。

以下に、アルゴリズムの詳細を記す。最初に、差分 DCT 係数を算出する。差分 DCT 係数を整数 N で除し、“0” になった係数をカウントする。このカウント値をマップ化したものをカウントマップと呼ぶ。原画像の任意のブロックにおけるカウント値がカウントマップの任意の座標における値となる。カウントマップの任意の座標 (x, y) において取りうるカウント値 $Val_{count}(x, y)$ の範囲は $0 \leq Val_{count}(x, y) \leq 63$ である。このカウント値 $Val_{count}(x, y)$ を閾値 Th_c で 2 値化し、“0” の場合を複雑ブロック、“1” の場合を平坦ブロックと分類する。

任意の着目ブロックが複雑ブロックであった場合、着目ブロックの近傍に一つでも平坦ブロックがある場合、このブロックをエッジ領域に属するとする。それ以外は複雑領域とする。ここで、2 値化閾値 $Th_c = 50$ とした。

平坦ブロックと判定されたブロックにおいて、近傍のブロックに平坦ブロックがあり、周囲の平坦ブロックの輝度平均を各々算出し、その差分絶対値が閾値 Th_{diff} を超えた場合にグラデーション領域とする。またそれ以外を平坦領域とする。ここで、ブロックにおける輝度平均は基本レイヤにおけるローカルデコード画像の値を用いる。

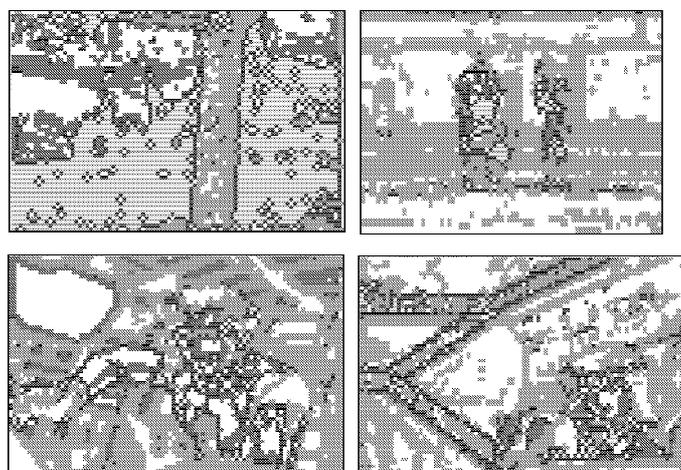


図 7.6: 画像特徴抽出実験結果．左上：Flower garden，右上：Baseball，左下：Robot1，右下：Robot2

グラデーション領域と判定されたブロックにおいて，輝度平均が Th_y を超える場合を高輝度グラデーション領域とする．また，それ以外を低輝度グラデーション領域とする．

このように，全てのブロックがエッジ領域，複雑領域，高輝度グラデーション領域，低輝度グラデーション領域，平坦領域に分類される．

なお，本アルゴリズムは，ピクチャタイプやマクロブロックのモードに拠らず，一元的に差分 DCT 係数から主観画質への寄与を推定するものである．

7.4.3 画像特徴抽出実験

前章で用いた 4 つの画像に画像特徴抽出アルゴリズムを適用し，画像特徴抽出実験を行った．実験で用いたパラメータを以下に示す．基本レイヤは 1 Mbps にてレート制御の下，符号化したものを使用した．

- カウントマップ作成のための除数 N : 8
- カウントマップの 2 値化閾値 Th_c : 50
- グラデーション領域判定のための閾値 Th_{diff} : 50
- グラデーション領域分割のための閾値 Th_y : 150

図 7.6 に実験結果を示す．ブロック毎に領域进行分类する．明度が低い順に，エッジ領域，高輝度グラデーション領域，低輝度グラデーション領域，複雑領域，平坦領域を示す．図 7.6 より，以下のことが読み取れる．”Flower Garden”においては，「空」と他のテクスチャとの境界にエッジ領域が出現し，「花畑」の部分は複雑領域と判定されている．

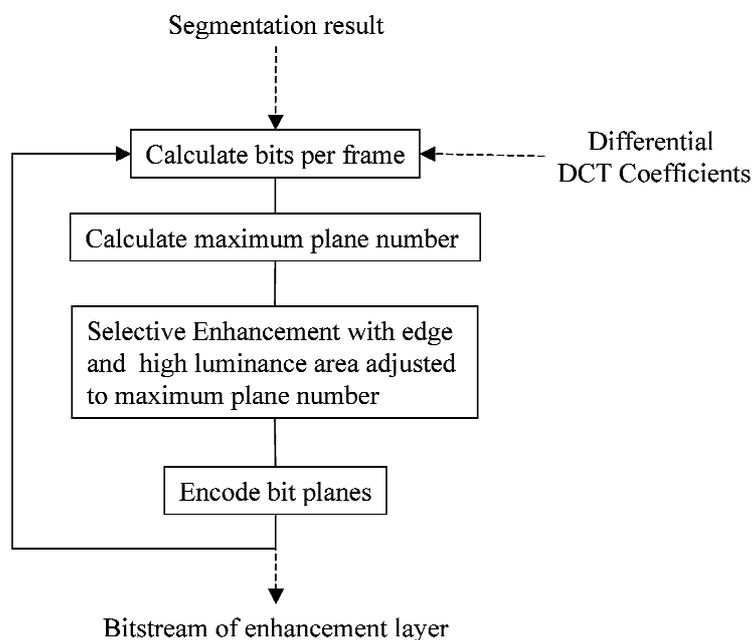


図 7.7: 提案 SE 方式 .

一般的なエッジ検出フィルタ³を用いて画像を処理するとエッジ部分が抽出されるが、「花畑」はもエッジとして抽出される。これに対して、本アルゴリズムでは、「花畑」は複雑領域と判断し、「花畑」と「空」の境界をエッジ領域と判断している。また、「Baseball」においては、バッター、審判、キャッチャーの輪郭部分がエッジ領域と判定され、背景部分はグラデーション領域として判定されている。「Robot1」、「Robot2」においても、線画の部分にエッジ領域が出現し、均等色領域には、高輝度および低輝度グラデーション領域と判定され、ほぼ、直感による判別と同程度の判定結果となった。

7.5 スケーラブル符号化実験

本節では、提案画像特徴抽出方法を符号化に適用してスケーラブルなビットストリームを作成し、その品質を PSNR および符号化画像のノイズの現れ方から評価する。

7.5.1 ビットプレーンシフトにおける基本方針

以下に本節におけるビットプレーンシフトに関する方針を示す。SE には、(1) どの MB を、(2) 何ビットシフトするか、という 2 つの自由度がある。本節では、シフトすべき MB とシフトす

³Sobel Filter など。

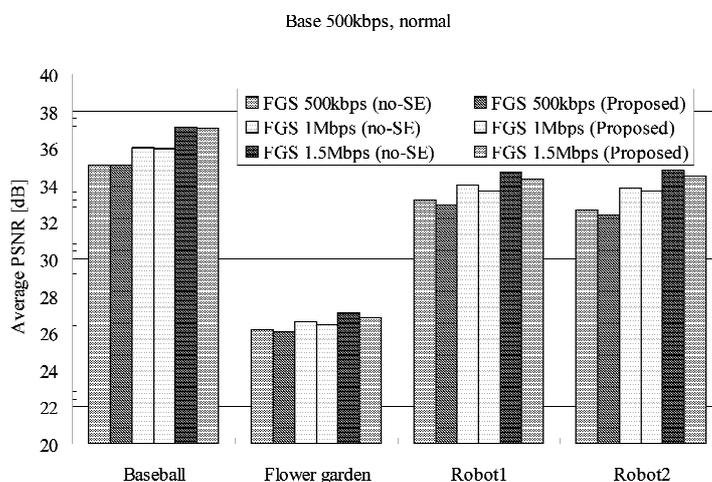


図 7.8: PSNR によるスケーラブル符号化結果 (通常スケーラブル符号化方式) .

るビット量を決定する。FGS 符号化の規格上，上位 4 プレーンのみが符号化される。図 7.3 で示したように，エッジ領域と複雑領域はビットプレーン数が多い。この 2 つの領域を極端にビットシフトすると，相対的に他の下位プレーンの符号化優先度が下がる。視覚上重要なデータが上位 4 プレーンに収まるように，ビットプレーンをシフトすることが望ましい。本節では，ブロックノイズが目立つ高輝度グラデーション領域と，モスキートノイズが目立つエッジ領域のビットプレーンをそのフレームにおける最大ビットプレーンまでシフトする方針とする。

図 7.7 にビットプレーンシフト方式を示す。最初に，拡張レイヤのビットレートから，1 フレーム当たりのビット数を算出する。算出されたビット数に達するまで，FGS 符号におけるビットプレーンを符号化する。差分 DCT 係数から，そのフレームにおける最大ビットプレーン値を算出する。本提案画像特徴抽出法にて領域分割された結果から，エッジ領域および高輝度グラデーション領域の各 MB における最大振幅値を最大ビットプレーン値にあわせてビットプレーンシフトする。

7.5.2 符号化実験

本提案画像特徴抽出アルゴリズムを図 7.1-(b) および (c) の符号化器に適用し，符号化実験を行い，画像品質を PSNR と主観画質から評価を行った。

便宜上，一般的な MPEG-4 スケーラブル符号化器構成である図 7.1-(b) を通常スケーラブル符号化方式 (Normal)，基本レイヤにおいて画像の縮小拡大を行う図 7.1-(c) を拡大縮小スケーラブル方式 (Scaling) と呼ぶ。2 つのスケーラブル符号化方法において，SE を用いない方式 (no-SE) と本画像特徴アルゴリズムとビットプレーンシフト方式を用いた方式 (提案 SE 方式, Proposed) の 2 つの方法を適用し，合計 4 つのパターンのスケーラブル符号化を比較検討する。符号化パラ

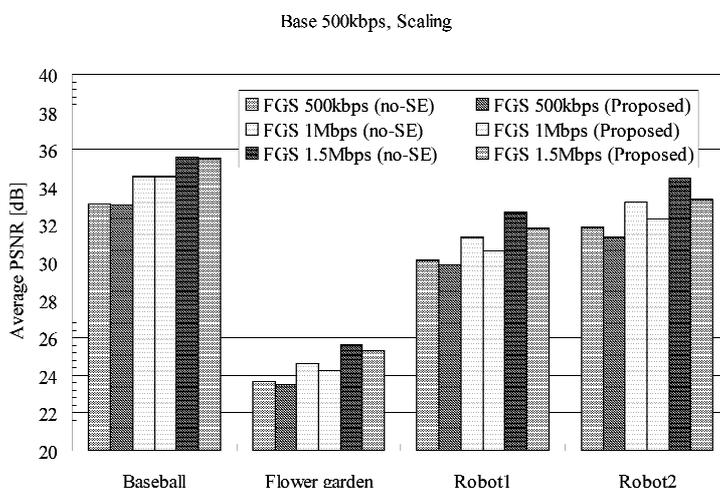


図 7.9: PSNR によるスケーラブル符号化結果 (拡大縮小スケーラブル符号化方式) .

メータは以下のとおりである .

- サイズ : VGA
- ビットレート :
 - 基本レイヤ : 500kbps, 拡張レイヤ : 500kbps
 - 基本レイヤ : 500kbps, 拡張レイヤ : 1Mbps
 - 基本レイヤ : 500kbps, 拡張レイヤ : 1.5Mbps

基本レイヤには, MPEG-4 ASP VM17.0 ベースの符号化方式を用いた . これには B ピクチャは含まれない . Inter/Intra MB 適応に関しては, ほぼ VM どれいのものを用いている . また, 特別な量子化適応や量子化制御は行っていない . 基本レイヤを 500kbps で符号化し, 拡張レイヤは, 500kbps, 1Mbps, 1.5Mbps の 3 通りのビットレートで符号化した . こうすることで, 基本レイヤ 500kbps, 拡張レイヤ 1.5Mbps, 合計 2Mbps で符号化した一つのビットストリームから 1Mbps, 1.5Mbps, 2Mbps の 3 種類のビットストリームを取り出すこと (One source multi use) と等価な評価が可能である . 符号化実験には前章と同じく "Flower garden", "Baseball", "Robot1", "Robot2" の 4 つの画像を用いる .

図 7.8, 図 7.9 に, 各画像をそれぞれ通常スケーラブル符号化方式および拡大縮小スケーラブル符号化方式にて符号化した時の, 原画像と符号化画像のシーケンスの平均 PSNR を示す .

最初に, 通常スケーラブル方式について, 本提案 SE 方式と SE を行わない方法について比較, 考察を行う . "Flower garden" において, 提案 SE 方式の顕著な品質の向上が見られた . 拡張レイヤを 500kbps および 1.5Mbps で符号化した画像を図 7.10 に示す .

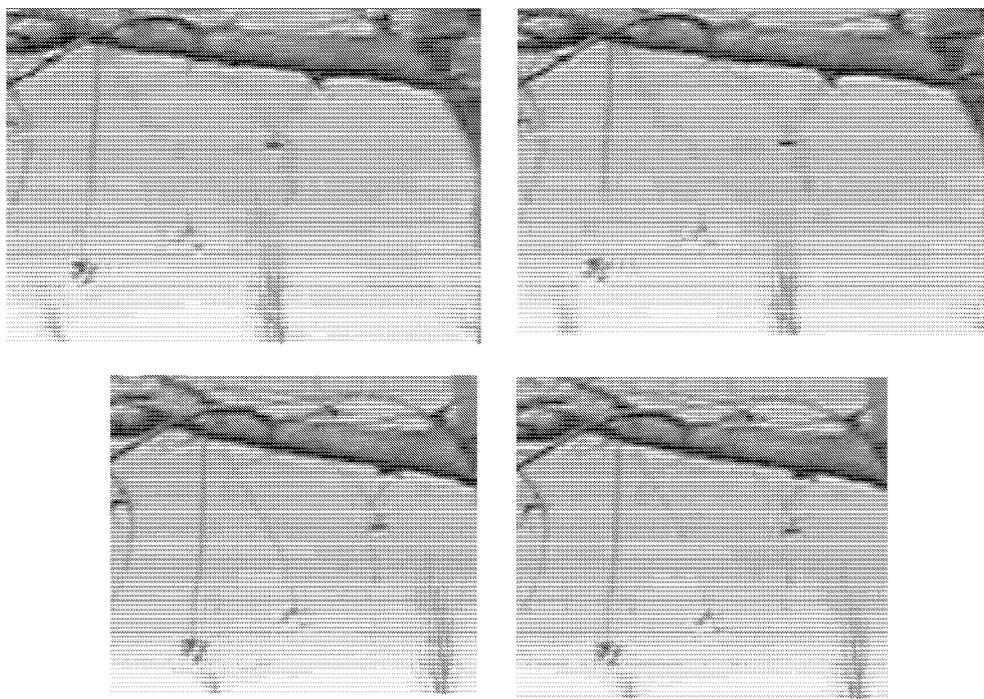


図 7.10: "Flower garden"における符号化画像の比較 (Normal) . 上段: 拡張レイヤ 500kbps における符号化画像 (左: 提案方式, 右: 従来方式, 以下同様), 下段: 拡張レイヤ 1.5Mbps における符号化画像 .

SEを行わない方法に比べて提案SE方式の「空」の領域のブロックノイズが軽減されていることが確認できる。また、SEを用いない場合では、「木」の一部が欠落したように見えるのに対し、提案SE方式ではつながって見える。複雑領域（「花畑」等）の差分DCT係数の絶対値が大きく、「空」の差分DCT係数が現れるのが相対的に低い順位のビットプレーンであるため、拡張レイヤを高いビットレートで符号化しても、十分に表現できるビットプレーンまで符号化することが困難であったためと考えられる。両方式の平均PSNRに着目すると、提案SE方式が0.3dB以下の劣化にとどまっている。また、筆者らが確認するところでは⁴、「Flower garden」において、相対的に符号量の割り当ての減少する複雑領域と判定された「花畑」の部分にて、従来法に比較して顕著な画質の劣化は見られなかった。「Robot1」においては、動きが激しいフレームにおいて、提案SE方式の顕著な品質の向上が見られた。拡張レイヤを1.5Mbpsで符号化した場合の様子を図7.11に示す。提案SE方式においてブロックノイズが軽減されていることが確認できる。平均PSNRに着目すると、提案SE方式がSEを行わない場合に比べて、0.25~0.4dB程度の劣化が認められた。「Baseball」および「Robot2」においては、両方式とも主観品質において顕著な差

⁴複数被験者による画像全体の主観評価は今後の課題としたい。

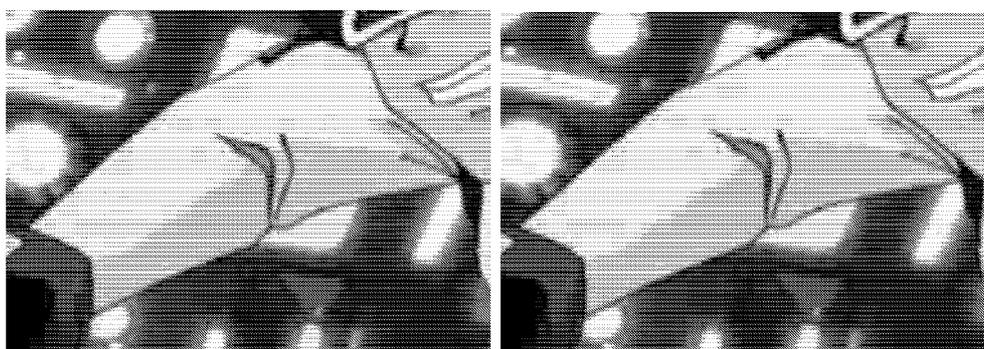


図 7.11: "Robot1"における符号化画像の比較 (Normal) . 拡張レイヤ 1.5Mbps における符号化画像 (左: 提案方式, 右: 従来方式) .

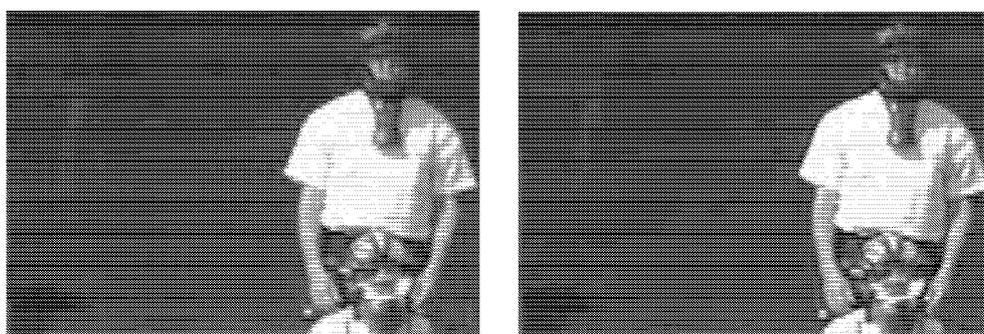


図 7.12: "Baseball"における符号化画像の比較 . 拡張レイヤ 500kbps における符号化画像 (左: 拡大縮小方式, 右: 通常方式) .

は見られなかった⁵ . "Baseball"における両方式の平均 PSNR の差がほとんどゼロであることから, 主観品質に差が見られないが, PSNR の劣化もほとんどないことが分かる . "Robot2"における, 提案 SE 方式の平均 PSNR は SE を用いない場合に比べて 0.25 ~ 0.3dB 程度の劣化にとどまった . 提案 SE 方式は, PSNR の向上には寄与しない高輝度グラデーション領域が優先されるように符号化しているので, PSNR の劣化が起こるのは必然である . しかし, PSNR の寄与にかかわる複雑領域とエッジ領域の優先順位を下げる作用はしないので, 極端な PSNR の劣化は回避できたと推測する .

次に, 提案 SE 方式における通常スケーラブル方式と拡大縮小スケーラブル方式の品質の差異について考察を行う .

"Baseball" および "Robot2" にて, 両スケーラブル方式に顕著な差異が見られた . 拡張レイヤを 500kbps で符号化した場合の様子を図 7.12 ("Baseball") および図 7.13 ("Robot2")⁶ に示

⁵紙面の関係上, 品質に差異のない画像の例は割愛する .

⁶Robot2 はパニングのある画像である . 図 6.4, 6.6 では, 領域分割が直感的に確認できる部分を, また, 図 1 3

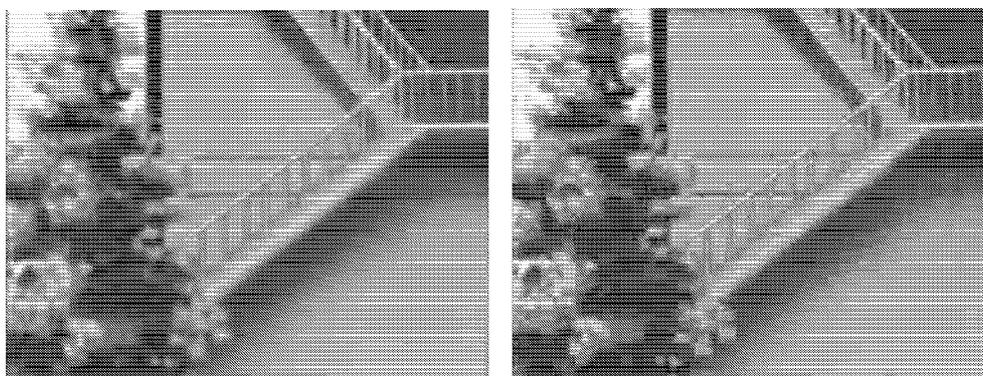


図 7.13: "Robot2"における符号化画像の比較．拡張レイヤ 500kbps における符号化画像（左：拡大縮小方式，右：通常方式）．

す．"baseball"においては「バックネット」の部分において、顕著な差が見られる．また，"Robot2"においては、「階段」部分のエッジの鮮明度と「壁」の部分のブロックノイズの現れ方において、拡大縮小方式に明らかな改善が見られた．ブロックノイズが低減された領域は、主に低輝度グラデーション領域である．この領域は、基本レイヤを縮小することで、少ない符号量で十分な品質を達成することができる．また、この領域は高周波成分を含まない領域であるので、縮小拡大処理に伴う高周波成分がカットされることによる主観品質の劣化を伴わない．エッジなどの高周波は拡張レイヤにて符号化されたため、品質が向上したと考えられる．一方で，"Flower garden"や"Robot1"では、高輝度グラデーション領域のブロックノイズの改善が見られたが、エッジ領域や複雑領域の品質劣化が見られた⁷．これは、両画像ともエッジ領域、複雑領域の占める割合が多い画像であり、加えて、基本レイヤが 500kbps と低レートであったことが品質の劣化を招いたと考えられる．すなわち、縮小拡大処理によって高周波が消去された画像から差分 DCT 係数をもとめるため、大きな振幅をもつ差分 DCT 係数の割合が高くなり、本実験に用いた条件下では十分な符号量が与えられなかったためと考えられる．

全画像において、PSNR は通常スケーラブル方式を用いた場合よりも、さらに 2dB 程度劣化する傾向があった．基本レイヤの縮小拡大処理が効果をもたらすのは、エッジ領域や複雑領域が十分少なく、グラデーション領域の面積が比較的多い画像であると言える．反対に、画像品質の劣化をもたらすのは、グラデーション領域に比べてエッジ領域や複雑領域の多い画像であると言える．よって、全般的に複雑な画像は、基本レイヤにて十分な符号量があった場合でも、解像度が半分になるため、拡張レイヤにてこれを補うに十分な符号量を割り当てない場合は、品質の高い画像は得られないと予想できる．

One Source Multi Use というアプリケーションの観点から、本節では低いビットレートにてビットストリームがカットされた場合でも、高い主観品質を実現することを目的とした検討を行っ

では、提案アルゴリズムの効果の顕著な部分をセレクトして示した．

⁷画像は紙面の都合上割愛する．

た．一般的に主観画質の劣化要因であると考えられている2つのノイズが現れる領域において，拡張レイヤの符号化データが優先的に符号化されるアルゴリズムを提案し，同一ビットレートにおける符号化画像の比較を行った．符号化画像における主観品質に寄与するノイズの現れ方から，提案 SE 手法は SE を用いない方法に比べて，視覚的に良好な品質を達成した．また，提案 SE 法によって，相対的に割り当て符号量が減少する部分（例えば，本節では”Flower garden”における「花畑」）において，特に目立った画質の劣化は見られなかった．よって，提案 SE 法は SE を用いない方法に比べて，低いビットレートでも主観品質を保持して符号化できる見通しを得たと言える．また，画像によっては，基本レイヤを一度縮小して符号化し，符号化画像を拡大して原画像と差分をとる場合が，主観品質に顕著な向上をもたらす場合があることを示した．画像全体の主観評価を交えた考察は今後の課題としたい．

7.6 本章のまとめ

本章では，差分 DCT 係数の性質から，画像の主観品質を左右する2つの歪であるブロックノイズおよびモスキートノイズが出現するブロックを差分 DCT 係数およびローカルデコード画像から推定する画像特徴抽出法を提案した．また，本画像特徴抽出法によって領域分割された結果から，ビットプレーンシフトを行う領域の選択と，ビットシフト量を決定するビットプレーンシフトアルゴリズムを提案した．また，本提案手法をスケーラブル符号化に適用し，主観品質に寄与するブロックノイズとモスキートノイズの現れ方から考察を行った．その結果，One source multi use にて，一つのビットストリームから SE を行わない方法に比べて高い主観品質のビットストリームを取り出せることを示した．

第8章 結論

本論文では、映像特徴を定義し、映像特徴の抽出法を提案し、それを映像符号化法に適用し、従来符号化方式である MC+DCT 方式と比較して、劇的な圧縮効率を達成可能であることを示した。本研究では、スプライト符号化を用いたオブジェクト符号化型アプローチならびに、FGS 符号化を用いた階層符号化型アプローチを用い、それぞれ一般的に画像に適用できることを示した。

第1章、第2章では、研究の背景、動機、および本研究のアプローチについて述べた。インターネットやモバイル通信環境がブロードバンド化しつつも、依然として 1Mbps 以下の中低レートの帯域での通信が中心であることを示し、効率よく映像を圧縮する方法として、映像特徴に適応的に符号化することを提案した。また、映像特徴を適応的に用いる映像符号化方法について、(1) パラメトリック符号化型、(2) オブジェクト符号化型、(3) 階層符号化型の3種類のアプローチに分類し、各々のアプローチについて、従来技術における映像特徴の定義、抽出法、符号化への適用、および問題点について述べた。また、本論文はオブジェクト符号化型と階層符号化型の2つのアプローチを軸に、新たな映像特徴の抽出法と画像符号化への適用法に関する研究を行うことを述べた。

第3章では、映像特徴の一つであるカメラ操作（ヘルマートモデル）と局所的な動き（平行移動モデル）の関係を定式化した。その関係から大局的な動きであるカメラパラメータを算出するアルゴリズムを提案した。さらに、それを基に各カメラ操作があった場合のスプライトを自動的に生成するアルゴリズムを提案した。さらに、実際に短い映像を適用し、実際のカメラモーションに近いパラメータを抽出でき、カメラモーションのみを反映した背景画像をスプライトとして良好に抽出できることを示した。しかし、本章の報告では隣接フレーム間カメラパラメータを輝度値を用いて検定を行っていないので、平坦な色使いの画像の場合、動きベクトルがゼロの値に集中することが予測され、隣接フレーム間カメラパラメータが誤抽出されることが考えられる。そこで、隣接フレーム間カメラパラメータ算出時における何らかの検定法と、絶対カメラパラメータ算出時の誤差の吸収法の確立が望まれること、完全にアウトライヤの影響を除去したカメラパラメータを算出するには前景となるオブジェクトを切り出し、背景部分のみでカメラパラメータを推定する必要があることを指摘した。

第4章では、背景映像のスプライト符号化への適用を目的とし、スプライト生成に適したグローバルモーション算出法の改良を行った。具体的には、輝度勾配対策とクラスタ選択を特徴としたアウトライヤの影響の少ないグローバルモーション算出アルゴリズムを提案し、算出精度の評価を行なった。輝度勾配対策は、アニメーション画像において効果を発揮した。クラスタ選択はカメラの動きのみを反映したグローバルモーション算出においてその精度向上に大きく寄与することを示した。また、画像中の背景部分（背景画像）に対して、実際にグローバルモーション

を用いた数種類の符号化法方式 (1)LME 符号化 (従来符号化パラダイムである MC+DCT ベース), (2)GMC 符号化, (3) ダイナミックスプライト符号化, (4) スタティックスプライト符号化を行った。現状では再現の忠実性に課題を残すものの, スタティックスプライトはその他の符号化方法に比べて高い圧縮効率を達成出来ることが分かった。今後はスプライト生成の高精度化, Low-latency スプライトの検討を課題とした。

第5章では, スプライト符号化の課題の一つであるビデオオブジェクトの自動抽出アルゴリズムを提案した。また, 提案アルゴリズムを MPEG-4 スプライト符号化に適用して MPEG-4 通常符号化方法と比較, 検討を行った。前景比率が画像全体の 10-15 %程度に収まる条件下で, MPEG-4 通常符号化方法の 1/4-1/2 以下の符号量で同程度の画質を実現した。また, 実際に超低レート (128kbps, 64kbps) で符号化実験し, 同程度の SNR においてスプライトモードがノーマルモードの倍以上のフレームレートを達成できることを示した。このように, スプライト符号化は特定の条件の下で, 劇的な符号量削減が期待できる。しかし, すべてのビデオシーケンスがこの条件を満たしているわけではなく, スプライト符号化できるショットを判定する方法の検討が必要であることが分かった。また, 実際に配信を考慮した場合のビデオオブジェクト毎の符号量割り当ての検討を今後の研究課題とした。

第6章では, スプライトモードが適用できるショットを自動的に判定する方法を取り入れ, スプライトモードとノーマルモードを適応的に切り替えて, すべての映像に適用できる汎用符号化法「マルチモード符号化方式」を提案し, スプライト符号化を用いた汎用的な MPEG-4 動画像符号化システムを開発したことを報告した。スプライトモードが優位なショットを自動的に抽出し, そのショットにはスプライトモードを適用し, その他のショットに対しては従来符号化法を踏襲するノーマルモードを適用するマルチモード符号化法を提案した。さらに, AVI データ等を逐次的に処理するための逐次処理方法を提案した。また, グローバルな動きの算出法, 前景領域抽出法, スプライト生成法に改良を行った。また, 前景背景に自動的に符号量を配分するためのマルチビデオオブジェクトレート制御法を提案した。また, 実際に符号化実験を行って, 低レートにおける従来符号化方式に対する本システムの優位性を示した。一方で, レートが高くなると従来符号化方式でも十分な映像品質が得られることを実験的に導き出し, 本システムを適用した場合でも劇的な圧縮効果をもたらすビットレートには限界があることを述べた。

第7章では, 階層符号化型のアプローチから, 画像特徴を用いた映像符号化の検討を行った。FGS 符号化における Selective Enhancement 機能を用いて基本レイヤにおいて主観品質が悪化する部分の符号量を補填して, トータルの画像品質を向上させる方法の検討を行った。FGS 符号化において, 基本レイヤの量子化前の DCT 係数と量子化後の DCT 係数の差分情報が拡張レイヤとして FGS 符号化されることを導出し, 実証実験を行った。また, この基本レイヤの量子化前の DCT 係数の性質を利用し, 画像の主観品質を左右する 2 つの歪であるブロックノイズおよびモスキートノイズが出現するブロックを DCT 係数およびローカルデコード画像から推定する方法を提案した。また, 本提案手法にて抽出された主観品質に寄与する領域を拡張レイヤにて Fine Granularity Scalable (FGS) 符号化することで, 主観品質を向上させることを実証した。

本研究のベースとなっている, スプライト符号化および FGS 符号化であるが, 次世代の三次元符号化 (3DAV Coding) およびスケーラブル符号化 (MPEG-21 Scalable Video Coding) のツ-

ルとして提案される動きがある。これらの次世代符号化方式において、グローバルモーション算出、スプライト生成、画質向上などの要求において、本研究は良好な現実解をあたえていくものと思われる。

謝辞

本論文をまとめるにあたって、ご指導いただいた、日本電信電話株式会社 OB で SC29 議長でもある早稲田大学国際情報通信研究科の渡辺裕教授に深く感謝いたします。また、筆者が早稲田大学理工学研究科修士課程在学時の恩師で日本電信電公社 OB でもある早稲田大学国際情報研究センターの富永英義教授に深く感謝いたします。本研究を遂行するにあたって貴重な討論をいただいた早稲田大学国際情報通信研究科の亀山渉教授，早稲田大学富永研究室の皆様，早稲田大学国際情報通信研究科渡辺研究室の皆様，日本電信電話株式会社サイバースペース研究所画像メディア通信プロジェクト映像符号化技術グループの皆様に厚くお礼を申し上げます。

2005 年 2 月 秦泉寺 久美

参考文献

- [1] ”平成 15 年通信利用動向調査の結果”，総務省 .
- [2] ”インターネット白書 2004”，インプレス社 .
- [3] ”Information technology- Coding of audio-visual objects-Part2:Visual Amendment 1:Visual extensions ISO/IEC 14496-2
- [4] ”AMENDMENT 4: Streaming video profile,” ISO/IEC 14496-2:1999/FDAM4
- [5] 斎藤隆弘，相澤清晴，“ビデオデータ圧縮の新しい展開”，”テレビジョン学会誌 Vol.49, No.4, pp.523-534, 1995.
- [6] 青木啓史，宮崎隆，“ソフトウェア MPEG-4 ASP ビデオコーデックの開発”，”情報処理学会第 39 回 AVM 研究会 39-9, pp.45-50.
- [7] 米山暁夫，柳原広昌，笹野義二，中島康之，“画像特徴に応じた MPEG-4 ビデオ符号化方式”，”情報処理学会第 37 回 AVM 研究会 37-2, pp.9-14.
- [8] James R. Bergen, P. Anandan, Keith J. Hanna, Rajesh Hingorani, ”Hierarchical Model-Based Motion Estimation,” European Conference on Computer Vision(ECCV) '92, pp.237-252, May 1992.
- [9] Michal Irani, P. Anandan, Jim Bergen, Rakesh Kumar, Steve Hsu, ”Efficient Representations of Video Sequences and Their Applications,” SignalProcessing: Image Communication Vol.8 pp.327-351, 1996.
- [10] Michal Irani, Steve Hsu, P. Anandan, ”Video Compression Using Mosaic Representations,” SignalProcessing: Image Communication Vol.7 pp.529-552, 1995.
- [11] John Y.A. Wang, Edward H. Adelson, Ujjaval Desai, ”Applying Mid-level Vision Techniques for Video Data Compression and Manipulation,” SPIE Vol. 2187 pp.116-127.
- [12] John Y.A. Wang, Edward H. Adelson, ”Representing Moving Images with Layers,” IEEE Trans. on Circuits and Systems for video Technology, Vol. 3, No. 5, September 1994.

- [13] 上倉一人, 渡辺裕, “動画像符号化におけるグローバル動き補償法,” 信学会論文誌 B-I Vol.J76-B-I No.12, pp.944-952, 1993 年 12 月 .
- [14] Kazuto Kamikura, Hiroshi Watanabe, “VideoCoding for Digital Strage Media Using Hierarchical Intraframe Scheme,” SPIE Vol.1360 Visual Communication And Image Processing 1990.
- [15] Aljoscha Smolic, Thomas Sikora, Jens-Rainer Ohm, “Long-Term Global Motion Estimation and Itsu Application for Sprite Coding, Content Description, and Segmentation,” IEEE Trans. on Circuit And Systems for Video Technology, Vol. 9, No. 8, December 1999.
- [16] Aljscha Smolic, Karsten Muller, Jens-Rainer Ohm, “Global Motion Compensation and Video Mosaicing Using Different 2-D Motion Models,” Picture Coding Symposium2001(PCS2001), pp.331-334, April 2001.
- [17] 如澤裕尚, “アフィン変換を用いた動き補償予測に関する検討,” 信学技報 IE94-36, 1994 年 7 月 .
- [18] Hirohisa Jozawa, “Segmentation-based Video Coding Using An Affine Motion Model,” SPIE Vol.2308, Visual Communication and Image Processing’94 September 1994.
- [19] Hirohisa Jozawa, Kazuto Kamikura, Atsushi Sagata, Hiroshi Kotera, Hiroshi Watanabe, “Two-Stage Motion Compensation Using Adaptive Global MC and Local Affine MC,” IEEE Trans. on Curcuits and Systems for Video Technology, Vol. 7, No. 1, pp.75-85, Febbruary 1997.
- [20] M. Kass, A. Witikin and D. Terzopoulos, “SNAKES: Active Contour Modeis,” Proceedings if 1st ICCV, pp279-286, APril 1997.
- [21] Philippe Salembier, “Morphological Multiscale Segmentation for Image Coding,” Signal Processing Vol.38, 00.359-386, 1994.
- [22] Philippe Salembier, Montse Pargas, “Hierachical Morphological Segmentation for Image Sequence Coding,” IEEE Trans. on Image Processing, Vol.3, No.5, pp.639-651, September 1994.
- [23] Philippe Salembier, Patrick Brigger, Josep R. Casas, Montse Pargas, “Morphological Operators for Image and Video Compression,” IEEE Trans. on Image Processing, Vol.5, No.6, pp.881-898, June 1996.
- [24] Philippe Salembier, Luis Torres, Fernand Meyer, Chuang Gu, “Region-Based Video Coding Using Mathematical Morphology,” Proceedings of The IEEE, Vol.83, No.6, pp.843-857, June 1995.

- [25] 境田慎一, 金次保明, “背景差分法と時空間 watershed 空間による領域成長法を併用した動画像オブジェクトの抽出,” 信学技報 IE99-151, pp.15-22, 2000年3月.
- [26] 酒井洋, 滝田久美, 中村能章, “動領域抽出に関する一方法,” テレビジョン学会技術報告 Vol.16, No.32, pp.19-23, 1992年6月.
- [27] Jae Gark Choi, Si-Woong Lee, Seong-Dae Kim, “Spatio-Temporal Video Segmentation Using a Joint Similarity Measure,” IEEE Trans. on Circuits And Systems fir Video Technology, Vol.7, No.2, pp.279-286, April 1997.
- [28] A. Neri, S. Colonnese, G. Russo, “Automatic Moving Object and Background Segmentation by Means of Higher order Statistics,” SPIE Vol.3024, VisualCommunication and Image Processing, pp.246-256, 1997.
- [29] A. Neri, S. Colonnese, G. Russo, “Video Sequence Segmentation for Object-based Coders Using Higher Order Statistics,” IEEE International Symposium on Circuits and Systems, pp.1245-1248, June 1997.
- [30] Lee M. Garth, H. Vincent Poor, “Detection of Non-Gaussian Signals: A Paradigm for Modern Statistical Signal Processing,” Proceedings of THE IEEE, Vol. 82, No.7, pp.1061-1095, July 1994.
- [31] Til Aach, Andre Kaup, “Statistical Model-based Change Detection in Moving Video,” Signal Processing Vol. 31, pp.165-180, 1993.
- [32] Roland Mech, Michael Wollborn, “Noize Robust Method for Segmentation of Moving Objects in Video Sequences,” IEEE International Conference on Acoustic, Speech and Signal, pp.2657-2660, April 1997.
- [33] Luc Vincent, Pierre Soille, “Watersheds in Digital Spaces: An efficient Algorithm Based on Immersion Simulations,” IEEE Trans. on Patter Analysys and Machine Intelligence, Vol.13, No.6, pp.583-598, June 1991.
- [34] Laura Teodosio and Walter Bender, “Salient Video Stills: Content and Context Preserved,” ACM Multimedia 93, pp.39-45, 1993.
- [35] Ming-Chieh Lee, Wei-ge Chen, Chih-lung Bruce Lin, Chuang Gu, Tomislav Markoc, Steven I. Zabinsky, Ricard Szeliski, “A Layerd Video Object Coding System Using Sprite and Affine Motion Model,” IEEE Trans. on Curcuits and Systems for Video Technology, Vol. 7, No. 1, pp.-130-145, Febbruary 1997.
- [36] 上田剛弘, 金子正秀, 斎藤隆弘, 原島博, “複数フレームのマップを用いた平年記述による動画像の低ビットレート符号化,” 信学会論文誌 D-II Vol.J80-D-II No.12 pp.3117-3126 1997年12月.

- [37] 阿久津明人, 外村佳伸, 橋本秀雄, 大庭有二, “ 動画像インデクシングを目的としたカメラ操作の規定法, ” 信学会論文誌 D-II Vol.J75-D-II No.2 pp.226-235, 1992 年 2 月 .
- [38] A. Akutsu, Y. Tonomura, H. Hamada, ”VIDEOSTYLER: Multi-dimensional Video Computing for Eloquent Mdeia Interface,” International Conference on Image Processing(ICIP)'95, 1995.
- [39] 阿久津明人, 外村佳伸, “ 投影法を用いた映像の解析方法と映像ハンドリングへの応用, ” 信学会論文誌 D-II Vol.J79-D-II No.5, pp.675-686, 1996 年 5 月 .
- [40] 谷口行信, 阿久津明人, 外村佳伸, “ PanoramaEcerpts: パノラマ画像の自動生成・レイアウトによる映像一覧, ” 信学会論文誌 D-II Vol.J82-D-II No.3 pp.390-398 1999 年 3 月 .
- [41] 星野准一, 増田功, “ 観測・推定画像間の誤差最小化によるパノラマ動画像の生成, ” 信学会論文誌 D-II, Vol. J81-D-II, No.11, pp.2565-2573, 1998 年 11 月 .
- [42] 竹内俊一, 四分一大助, 寺島信義, 富永英義, “ ズームイン・ズームアウトを繰り返すビデオ画像シーケンスを用いた平面静止画の領域適応高精細化デジタル化, ” 信学会論文誌 D-II Vol.J83-D-II, No.12, pp.2675-2685, 2000 年 12 月 .
- [43] 竹内俊一, 寺島信義, 富永英義, “ 投影特徴を利用した階層的画像モザイク化手法による絵画等静止画像の高精細デジタル化, ” 信学会論文誌 D-II, Vol. J83-D-II, No4, pp.1090-1099, 2000 年 4 月 .
- [44] 田邊集, 秦泉寺久美, 渡辺裕, 富永英義, “MPEG-2 bitstream からの同領域抽出法の検討, ” 情報処理学会第 36 回 AVM 研究会 36-7, pp.37-42.
- [45] 中嶋淳一, 上倉一人, 八島由幸, “MPEG-4 FGS レイヤにおける符号化マクロブロック選択に関する考察, ” 情報処理学会第 40 回 AVM 研究会, pp.1-6.
- [46] Kumi Jinzenji, Satoshi Ishibashi, and Hiroshi kotera, ”Layered Sprite for Video Compression by Detecting Camera Movement,” International Workshop on Coding Techniques for Very Low Bit-rate Video (VLBV97), pp.9-12, July 1997.
- [47] Kumi Jinzenji, Satoshi Ishibashi, and Hiroshi kotera, ”Algorithm for Automatically Producing Layered Sprites by Detecting Camera Movement,” IEEE International Conference on Image Processing '97 (ICIP97), pp.767-770, October 1997.
- [48] ”MPEG-4 Video Verification Model version 11.0,” ISO/IEC JTC1/SC29/WG11/N2172
- [49] 秦泉寺久美, 渡辺裕, 小林直樹, “スプライト生成のためのグローバルモーション算出法と符号化への適用, ”電子情報通信学会和文論文誌 D-II, Vol.J83-D-II, No.2, pp.535-544, February 2000.

- [50] Kumi Jinzenji, Hiroshi Watanabe, and Naoki Kobayashi, "Global Motion Estimation for Static Sprite Production and Its Application to Video Coding," IEEE International Workshop on Intelligent Signal Processing And Communication Systems '98 (ISPACS98), pp.328-332, November 1998.
- [51] 秦泉寺久美, 石橋聡, 小林直樹, "カメラモーション抽出による多層スプライト自動生成," 電子情報通信学会和文論文誌 D-II, Vol.J82-D-II, No.6, pp.1018-1030, June 1999.
- [52] Kumi Jinzenji, Seishi Takamura, Hiroshi Watanabe and Naoki Kobayashi, "Automatic VOP Production Scheme for Very Low Bit Rate Video Coding," Picture Coding Symposium '99 (PCS99), pp.299-302, April 1999.
- [53] Kumi Jinzenji, Shigeki Okada, Hiroshi Watanabe, Naoki Kobayashi, "Automatic Two-layer Video Object Plane Generation Scheme And Its Application to MPEG-4 Video Coding," The 2000 IEEE International Symposium on Circuits and Systems (ISCAS2000), pp.606-609, May 2000.
- [54] "MPEG-4 Video Verification Model version 15.0," ISO/IEC JTC1/SC29/WG11/N3093
- [55] Thomas Meier, King N. Ngan, "Automatic Segmentation of Moving Objects for Video Object Plane Generation," IEEE Trans. on Circuits And Systems for Video Technology, Vol. 8, No.5, pp.525-538, September 1998.
- [56] Michal Irani, Benny Rousso, Shmuel Peleg, "Detecting and Tracking Multiple Moving Objects Using Temporal Integration," European Conference on Computer Vision(ECCV)'93, 1993.
- [57] 秦泉寺久美, 渡辺裕, 岡田重樹, 小林直樹:「スプライト符号化を用いた MPEG-4 動画像超高压縮」, 電子情報通信学会和文論文誌 D-II, Vol. J84-D-II No.5 pp.758-768, 2001
- [58] Kumi Jinzenji, Hiroshi Watanabe, Shigeki Okada, Naoki Kobayashi, "MPEG-4 Very Low Bitrate Video Compression Using Sprite Coding," IEEE ICME'01, TA0.02.
- [59] Shigeki Okada, Kumi Jinzenji, Hiroshi Watanabe, Naoki Kobayashi, "An Approach to MPEG-4 Multi Mode Coding Using Sprite Coding," PCS2001 pp.421-424, March 2001.
- [60] Kumi Jinzenji, Shigeki Okada, Naoki Kobayashi, Hiroshi Watanabe, "MPEG-4 Very Low Bitrate Video Coding by Adaptively Utilizing Sprite," ICME2002.
- [61] Anthony Vetro, Huifang Sun, Yao Wang, "MPEG-4 Rate Control for Multiple Video Objects," IEEE Trans. on Circuits And Systems for Video Technology, Vol.9, No.1, pp.186-199, February 1999.
- [62] "MPEG-4 Video Verification Model version 17.0," ISO/IEC JTC1/SC29/WG11/N3515

- [63] Weiping Li, "Overview of Fine Granularity Scalable in MPEG-4 Video Standard," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 11, No. 3, pp. 301-317, March 2001.
- [64] 中神央二, 渡辺裕, 富永英義, "差分画像を利用したアニメーション映像からのオブジェクト抽出," 情報処理学会第40回AVM研究会, pp. 31-36, March 2003.

業績

類別	題名	掲載雑誌名	発表年月	連名者
1. 論文				
○ (1)	DCT係数を利用した主観品質に寄与する画像特徴の抽出とMPEG-4スケーラブル符号化への適用	電子情報通信学会和文論文誌D-II	2004年9月 掲載予定	秦泉寺久美 中嶋淳一 上倉一人 八島由幸
○ (2)	スプライトを適応的に利用した映像符号化システム	画像電子学会論文誌Vol.33 No.3 pp.333-343	2004年5月	秦泉寺久美 岡田重樹 八島由幸 渡辺裕
○ (3)	スプライト符号化を用いたMPEG-4動画画像超高压縮	電子情報通信学会和文論文誌D-II, Vol. J84-D-II No.5 pp.758-768	2001年5月	秦泉寺久美 渡辺裕 岡田重樹 小林直樹
○ (4)	スプライト生成のためのグローバルモーション算出法と符号化への適用	電子情報通信学会和文論文誌D-II, Vol.J83-D-II, No.2, pp.535-544	2000年2月	秦泉寺久美 渡辺裕 小林直樹
○ (5)	カメラモーション抽出による多層スプライト自動生成	電子情報通信学会和文論文誌D-II, Vol.J82-D-II, No.6, pp.1018-1030	1999年6月	秦泉寺久美 石橋聡 小林直樹
2. 国際会議論文				
○ (6)	MPEG-4 Very Low Bitrate Video Coding by Adaptively Utilizing Sprite	IEEE International Conference on Multimedia and Expo. 2002 (ICME2002)	2002年8月	Kumi Jinzenji Shigeki Okada Naoki Kobayashi Hiroshi Watanabe
○ (7)	MPEG-4 Very Low Bitrate Video Compression Using Sprite Coding	IEEE International Conference on Multimedia and Expo. 2001 (ICME2001)	2001年8月	Kumi Jinzenji Hiroshi Watanabe, Shigeki Okada Naoki Kobayashi
○ (8)	Automatic Two-layer Video Object Plane Generation Scheme And Its Application to MPEG-4 Video Coding	IEEE International Symposium on Circuits and Systems (ISCAS2000), pp.606-609	2000年5月	Kumi Jinzenji Shigeki Okada Hiroshi Watanabe Naoki Kobayashi
○ (9)	Automatic VOP Production Scheme for Very Low Bit Rate Video Coding	Picture Coding Symposium '99 (PCS99), pp.299-302	1999年4月	Kumi Jinzenji Seishi Takamura Hiroshi Watanabe Naoki Kobayashi
○ (10)	Global Motion Estimation for Static Sprite Production and Its Application to Video Coding	IEEE International Workshop on Intelligent Signal Processing And Communication Systems '98 (ISPACS98), pp.328-332	1998年10月	Kumi Jinzenji Hiroshi Watanabe Naoki Kobayashi

類別	題名	掲載雑誌名	発表年月	連名者
○(11)	Algorithm for Automatically Producing Layered Sprites by Detecting Camera Movement	IEEE International Conference on Image Processing '97 (ICIP97), pp.767-770	1997年10月	Kumi Jinzenji Satoshi Ishibashi Hiroshi kotera
○(12)	Layered Sprite for Video Compression by Detecting Camera Movement," International Workshop on Coding Techniques for Very Low Bit-rate Video (VLBV97), pp.9-12	International Workshop on Coding Techniques for Very Low Bit-rate Video (VLBV97), pp.9-12,	1997年7月	Kumi Jinzenji Satoshi Ishibashi Hiroshi kotera
○(13)	Graphical Video Representation for Scalability	The First International Workshop on New Media Technology, pp.29-32	1996年3月	Kumi Jinzenji Hisashi Kasahara
(14)	Pixel-based Extraction of Moving Objects for Sprite Coding	The 6 th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI2002), Proc. Vol. III, pp.195-200	2002年7月	Shigemitsu Watanabe Kumi Jinzenji Hiroshi Watanabe
(15)	Sprite Coding in Object-based Video Coding Standard: MPEG-4	The 5 th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI2001)	2001年7月	Hiroshi Watanabe Kumi Jinzenji
(16)	An Approach to MPEG-4 Multi Mode Coding Using Sprite Coding	Picture Coding Symposium 2001(PCS2001) pp.421-424	2001年3月	Shigeki Okada Kumi Jinzenji Hiroshi Watanabe Naoki Kobayashi
3. 国内学会				
(17)	Selective Enhancementと基本レイヤの縮小を利用したMPEG-4スケーラブル符号化に関する一検討	電子情報通信学会IE研究会 IE2003-125, pp.61-66	2003年12月	秦泉寺久美 中嶋淳一 上倉一人 八島由幸
(18)	DCT係数を利用した主観品質に寄与する画像特徴の抽出とMPEG-4スケーラブル符号化への適用	電子情報通信学会IE研究会 IE2003-63, pp.45-50	2003年9月	秦泉寺久美 中嶋淳一 上倉一人 八島由幸
(19)	スプライトを用いたMPEG-4超高圧縮符号化方式: VideoESPER	情報処理学会オーディオビジュアル復号処理研究会 36-8, pp.43-51	2002年3月	秦泉寺久美 岡田重樹 小林直樹 渡辺裕 米原紀子
(20)	MPEG-2 bitstreamからの動領域判定法の検討	情報処理学会オーディオビジュアル復号処理研究会	2002年3月	田邊集 秦泉寺久美 渡辺裕 富永英義
(21)	MPEG-2 bitstreamからのROI抽出とトランスコーダへの適用	画像電子学会第30回年次大会, No.9, pp.113-114, June 2002.	2002年6月	田邊集 秦泉寺久美 渡辺裕 富永英義
(22)	Watershed変換を用いた動物体抽出に関する検討	画像電子学会第30回年次大会, No.8, pp.111-112, June 2002.	2002年6月	渡辺成光 秦泉寺久美 渡辺裕
(23)	MPEG-2 bitstreamからの動領域判定法の検討	電子情報通信学会総合大会D-11- 96	2002年3月	田邊集 秦泉寺久美 渡辺裕 富永英義

類別	題名	掲載雑誌名	発表年月	連名者
(24)	Pixel-based Extraction of Moving Objects for Sprite Coding	電子情報通信学会全国大会, D-11-93, Mar. 2002	2002年3月	渡辺成光 秦泉寺久美 渡辺裕
(25)	グローバル動き補償符号化における条件付画素補填	情報処理学会オーディオビジュアル復号処理研究会	2001年6月	秦泉寺久美 阪谷徹 八島由幸 小林直樹
(26)	スプライト符号化における解像度変換処理に関する一検討	電子情報通信学会総合大会 P.45	2001年3月	岡田重樹 高村誠之 秦泉寺久美 阪谷徹 小林直樹
(27)	スプライト符号化を利用したMPEG-4マルチモード符号化方式の研究	画像符号化シンポジウム (PCSJ)2000 pp.91-92	2000年11月	岡田重樹 秦泉寺久美 渡辺裕 小林直樹
(28)	MPEG-4スプライト符号化を用いた超低レート動画圧縮	画像電子学会0-03-02 pp.7-13	2000年9月	秦泉寺久美 岡田重樹 渡辺裕 小林直樹
(29)	背景スプライトを用いた動画圧縮に関する一考察	電子情報通信学会総合大会D-11-52	2000年3月	秦泉寺久美 岡田重樹 渡辺裕 小林直樹
(30)	スプライトに適したショット判定方法	電子情報通信学会総合大会 D-11-53	2000年3月	岡田重樹 秦泉寺久美 渡辺裕 小林直樹
(31)	スプライトに適したショットの判定方法	情報処理学会オーディオビジュアル復号処理研究会AVM28-11, pp.61-66	2000年3月	岡田重樹 秦泉寺久美 渡辺裕 小林直樹
(32)	二層ビデオオブジェクト自動生成とMPEG-4符号化への適用	信学技報IE99-115, pp.49-54	1999年12月	秦泉寺久美 渡辺裕 小林直樹
(33)	VOP自動生成とMPEG-4符号化への適用	画像符号化シンポジウム, P-4.03, pp.73-74	1999年9月	秦泉寺久美 渡辺裕 小林直樹
(34)	フレーム間差分の累積を用いた動領域抽出法	映像情報メディア学会年次大会27-6, p.415	1998年7月	秦泉寺久美 渡辺裕 小林直樹
(35)	スタティックスプライト符号化における符号化効率の一検討	電子情報通信学会ソサエティ大会D-11-28	1998年9月	秦泉寺久美 渡辺裕 小林直樹
(36)	スタティックスプライト生成のためのグローバルモーション算出法と符号化への適用	信学技報IE-98-36	1998年7月	秦泉寺久美 渡辺裕 石橋聡 小林直樹

類別	題名	掲載雑誌名	発表年月	連名者
(37)	スプライト生成におけるアラインメントに関する一考察	信学技報IE-97-81	1997年10月	秦泉寺久美 石橋聡 小林直樹
(38)	スプライト生成におけるカメラモーション抽出法	電子情報通信学会ソサイエティ大会D-11-28	1997年9月	秦泉寺久美 石橋聡 小寺博
(39)	ショット映像圧縮のための多層スプライト自動生成	電子情報通信学会総合大会D-11-32	1997年3月	秦泉寺久美 石橋聡 小寺博
(40)	カメラ操作抽出によるスプライト生成と動画圧縮への利用	信学技報IE-96-139	1997年2月	秦泉寺久美 石橋聡 小寺博
(41)	遠隔プレゼンテーションにおける撮像画像からの資料切り出し	信学会春季大会	1994年3月	秦泉寺久美 加藤洋一
(42)	小領域内でのクラスタリングを用いた部品画像の半自動的抽出法	信学技報PRU93-23	1993年5月	秦泉寺久美 酒井洋 志和新一
(43)	小領域内でのクラスタリングを用いた部品画像の半自動的抽出法	信学技報PRU93-24	1993年6月	秦泉寺久美 酒井洋 志和新一
(44)	小ブロック内におけるクラスタリングを用いたインタラクティブな領域抽出法	信学会春季大会D-488	1993年3月	滝田(秦泉寺)久美 酒井洋 志和新一
4. 解説記事				
(45)	多視点映像からの実世界のモデリングと可視化	映像情報メディア学会誌, Vol.51, No.3, p335	1997年3月	秦泉寺久美
(46)	スプライトとアフィン動きモデルを用いた階層的ビデオオブジェクト符号化	映像情報メディア学会誌, Vol.51, No.6, pp.806-808	1997年6月	秦泉寺久美
(47)	モザイク表現を用いた映像圧縮	テレビジョン学会誌, Vol.50, No.10, pp.1624-1627	1996年10月	秦泉寺久美
(48)	画像圧縮のための多重解像度動き推定技術	テレビジョン学会誌, Vol.50, No.7, pp.984-985	1996年7月	秦泉寺久美
(49)	ビデオアルジャブラ(映像代数)を用いた映像の合成と検索	テレビジョン学会誌, Vol.49, No.12, p1636	1996年12月	秦泉寺久美
(50)	動画のコンテンツによる表現	テレビジョン学会誌, Vol.49, No.3 p403	1996年3月	秦泉寺久美
(51)	オールデジタルプロダクションセンタ	テレビジョン学会誌, Vol.49, No.7, p945	1995年7月	秦泉寺久美

図一覧

- 3.1 カメラモデル
- 3.2 スプライトモデル
- 3.3 平行移動動きベクトルの特徴空間での分布
- 3.4 スプライト生成アルゴリズム
- 3.5 ビデオクリップ内処理
- 3.6 絶対カメラパラメータ
- 3.7 原画像と予測画像の相関
- 3.8 生成されたスプライト”MIT sequence”
- 3.9 生成されたスプライト”Stefan”
- 3.10 生成されたスプライト”Flower garden”
- 3.11 背景の再生画像の例
- 3.12 背景の再生画像と原画像の相関 (MIT sequence)
- 4.1 GM 算出アルゴリズムのながれ
- 4.2 クラスタ選択法の概念
- 4.3 クラスタ選択法の詳細
- 4.4 GM 算出実験による提案手法の効果
- 4.5 スプライト生成による GM 算出の評価
- 4.6 復号画像の原画像に対する SN 比の変化
- 4.7 生成されたスタティックスプライト
- 4.8 復号画像の例

- 5.1 MPEG-4 符号化復号モデルと 2 層ビデオオブジェクト
- 5.2 2 層ビデオオブジェクト自動生成アルゴリズムの概略
- 5.3 XYT 時空間における画像のアラインメントの様子 . (a) 仮スプライト . (b) 本スプライト
- 5.4 前景オブジェクト形状のマクロブロック近似の様子 . (a) 元の形状 , (b) 一番粗い形状近似 , (c) 提案手法における初期マクロブロック化 , (d) 提案手法における拡張マクロブロック化
- 5.5 前景比率固定下における前景オブジェクト抽出アルゴリズム
- 5.6 前景比率と MPEG-4 スプライト符号化における符号量の関係 . MPEG-4 通常符号化時の符号量との比で表現
- 5.7 前景比率 10-15 % 程度 , 30fps における MPEG-4 スプライト符号化と MPEG-4 通常符号化の符号量の比較
- 5.8 $QP = 12$ 固定で符号化時の符号化画像の例 ("Horserace") . 左: ノーマルモード , 右: スプライトモード
- 5.9 背景スプライトの抽出例 . (a) Horserace , (b) Board , (c) Athlete
- 5.10 原画像 , 前景オブジェクト , 合成画像の例 . (a) Horserace , (b) Board , (c) Athlete
- 5.11 ノーマルモードおよびスプライトモードにおけるフレームレートと符号量の関係
- 5.12 同一ビットレート符号化時の SNR による画質評価 . 左 : 128kbps , 右 : 64kbps
- 5.13 128kbps におけるノーマルモードおよびスプライトモードの符号化画像の例 . 上段 : ノーマルモード , 下段 : スプライトモード
- 6.1 スプライトショットの概念
- 6.2 本提案符号化システムにおけるユニット処理の概念
- 6.3 本提案符号化システムのブロックダイアグラム
- 6.4 フレームにおける平均 QP の変化
- 6.5 符号化画像の例 . 左 , 提案方式 (本提案符号化システム) , 右 , 従来方式 (VM 17.0)
- 6.6 提案符号化システム専用デコーダ
- 6.7 スプライトビューワによる生成オブジェクトおよびグローバルモーションの例 . 上段左より , 原画像 , 符号化画像 , グローバルモーション . 下段左より , 前景オブジェクト , スプライトから切り出された背景画像

- 7.1 FGS を用いた MPEG-4 スケーラブル符号化器構成：(a) 一般的な構成（その 1）, (b) 一般的な構成（その 2）, (c) 基本レイヤに拡大・縮小処理を加えた構成
- 7.2 MPEG-4 FGS 符号化における Selective Enhancement の概念
- 7.3 基本レイヤにおける差分 DCT 係数の分布 (P ピクチャ)
- 7.4 差分 DCT 係数による領域分割例 . 上段より , ”Flower Garden” , ”Baseball” , ”Robot1” , ”Robot2”
- 7.5 差分 DCT 係数を用いた画像特徴抽出アルゴリズム
- 7.6 画像特徴抽出実験結果 . 左上 : Flower garden , 右上 : Baseball , 左下 : Robot1 , 右下 : Robot2
- 7.7 提案 SE 方式
- 7.8 PSNR によるスケーラブル符号化結果 (通常スケーラブル符号化方式)
- 7.9 PSNR によるスケーラブル符号化結果 (拡大縮小スケーラブル符号化方式)
- 7.10 ”Flower garden” における符号化画像の比較 (Normal) . 上段 : 拡張レイヤ 500kbps における符号化画像 (左 : 提案方式 , 右 : 従来方式 , 以下同様) , 下段 : 拡張レイヤ 1.5Mbps における符号化画像
- 7.11 ”Robot1” における符号化画像の比較 (Normal) . 拡張レイヤ 1.5Mbps における符号化画像 (左 : 提案方式 , 右 : 従来方式)
- 7.12 ”Baseball” における符号化画像の比較 . 拡張レイヤ 500kbps における符号化画像 (左 : 拡大縮小方式 , 右 : 通常方式)
- 7.13 ”Robot2” における符号化画像の比較 . 拡張レイヤ 500kbps における符号化画像 (左 : 拡大縮小方式 , 右 : 通常方式)

表一覧

- 1.1 画像特徴を用いた符号化のアプローチ
- 3.1 カメラパラメータの分類
- 3.2 使用した画像 (1)
- 3.3 使用した画像 (2)
- 3.4 各ビデオクリップにおけるカメラ操作
- 4.1GM 算出実験に用いた画像
- 4.2 画像符号化実験に用いた画像
- 4.3 符号化効率ならびに画品質
- 5.1MPEG-4 スプライト符号化実験に用いた画像
- 6.1 実験画像
- 6.2 スプライトショット抽出結果の一例 (”skateboard”)
- 6.3 平均前景率, スプライトショット符号化選択率
- 6.4 符号化結果