

論文 / 著書情報
Article / Book Information

題目(和文)	操作量に規範を設けたインタラクティブ知的符号化
Title(English)	
著者(和文)	大関和夫
Author(English)	
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:乙第3266号, 授与年月日:1999年2月28日, 学位の種別:論文博士, 審査員:
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:乙第3266号, Conferred date:1999/2/28, Degree Type:Thesis doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

操作量に規範を設けたインタラクティブ知的符号化

平成10年度

大 関 和 夫

目次

第1章 序論.....	1
1.1 研究の背景.....	1
1.2 研究の目的	5
1.3 インタラクティブ知的符号化方式.....	6
・適応KL変換符号化方式	7
1.4 本論文の構成	9
第2章 適応 KL 変換符号化方式	11
2.1 まえがき	11
2.2 従来のKL変換符号化研究について	12
2.3 適応適応可変行列 KL 変換符号化方式	14
2.4 SVD符号化方式との差異の検討	19
2.5 変換行列の設計	22
2.5.1 自己相関関数行列の距離による設計	22
2.5.2 変換誤差評価基準による設計	25
(可分量子化での設計)	
2.5.3 変換誤差評価基準による設計	26
(非可分量子化での設計)	
2.6 計算機による実験.....	27

目次

2.7 むすび	34
第3章 インタラクティブ知的符号化の基本構成.....	36
3.1 まえがき.....	36
3.2 画像の構造化と高能率符号化	36
3.3 次世代符号化方式の動向	41
3.4 インタラクティブ知的符号化の提案	43
3.4.1 インタラクティブ知的符号化の基本構成	43
3.4.2 インタラクティブ性の意義	45
3.5 むすび	47
第4章 グローバル動き検出とノイズ除去フィルタ	49
4.1 まえがき	49
4.2 パラメータの検出のシステム構成	49
4.3 3次元パラメータの検出	54
4.3.1 奥行き z の検出	55
4.3.2 3次元グローバル動きパラメータの検出	57
4.3.3 近似を行なう場合	58

目次

4.4 グローバル動きパラメータ検出の基本実験	61
4.5 ノイズ除去フィルタと検出精度	73
4.5.1 インタラクティブ操作と入力ノイズ	73
4.5.2 最小 2 乗法の適用条件	77
4.5.3 入力ノイズとLPF	82
4.5.4 仮想不動点の作用	83
4.5.5 構造型ノイズ除去フィルタ	85
4.6 特徴点の奥行き z の検出における精度評価	89
4.7 実験と考察	91
4.8 むすび	102
第 5 章 ローカル動き検出、表情、眼鏡画像処理	104
5.1 まえがき	104
5.2 ローカル動き検出と表情の記述	105
5.2.1 インタラクティブ知的符号化と問題点	105
5.2.2 ローカル動きパラメータ記述の操作量	109
5.2.3 周期パターンを用いた連動処理	113
5.2.4 連動規則の登録方式	117

目次

5.2.5 連動と操作量	117
5.3 操作量を考慮した眼鏡画像処理	126
5.3.1 眼鏡処理を行なわない場合の問題点	126
5.3.2 眼鏡画像処理	126
5.3.3 消え残りの評価	131
5.4 むすび	136
第6章 顔動画像の伝送実験と画質評価	137
6.1 まえがき	137
6.2 ローカル動きの記述方式による違い	137
6.3 再生画像	140
6.4 ヒューマンエージェントを目指した顔動画像伝送	158
6.4.1 顔動画像の伝送実験の構成	158
6.4.2 顔動画像の伝送実験の結果	163
6.4.3 ヒューマンエージェントとしての顔動画像符号化実験	166
6.5 主観評価実験	172
6.6 ビット量の評価	172

目次

6.7 むすび	174
第7章 結論	176
謝辞	183
参考文献	185
発表論文リスト	191
登録特許リスト	202

第1章 序論

1.1 研究の背景

画像の帯域圧縮技術は、伝送や記録をいかにして効率的に行い、それをいかにして源情報に対して忠実に行うかという課題を研究するため発展してきた[1-5]。デジタル画像の高効率符号化に対する基盤となる理論として、シャノンの情報理論、レート歪み理論、数学的には線形代数における2次形式や固有値問題などがある。1次元信号として、シャノンの情報理論を直接的に発展させた可逆な符号化方式としてハフマン符号化がある。これは、符号化事象の発生確率という統計量から、そのエントロピーに近い符号化性能を達成できる符号を生成できるものである。また、ハフマン符号は符号化操作に区切りをつけ、前後に関係を持たないブロック符号であるが、Elias 符号から発展した逐次処理を行なう算術符号は、エントロピーにより柔軟に近づけやすいなどの特徴を備えている。非可逆な符号化方式としては、DPCM (Differential Pulse Code Modulation) や変換符号化が代表的な方式である。DPCM は差分信号の分布の解析を基礎とし、これに分布の量子化の問題が加わっている。量子化に関しては Max の最適量子化やベクトル量子化に関するいくつかの知見が得られている。画像を2次元的な広がりを持つブロックとして扱うことによって、より高効率になる可能性が生じる。2次元画像ブロックの周波数分析において、主に低周波成分へのエネルギーの集中があることに注目した変換符号化がある。変換符号化ではフーリエ変換、アダマール変換、スラント変換、離散コサイン変換などが研究されてきた。これには、線形代数における2次形式や固有値問題などが基盤となっている。更に時間方向に第3の次元をとり、動き補償などの3次元信号処理を加えた動画画像符号化に対しては、3次元の周波数解析などに加え、前フレームと現在のフレームのパターンマッチングを行なうことが効果的であり、多数の研究がなされてきた。このように情報理論をはじめとする基盤的で有効な理論を基にした上で、現実には多岐にわたるデジタル画像の高効率な各種の符号化に対しては、数多くの個別のデジタル画像符号化処理方式の研究・開発がなされてきた。

一方、デジタル画像信号というメディアはファクシミリのような2値信号に始まり、静止自然画像、テレビジョン放送用動画画像、テレビ電話・会議用の通信用動画画像、蓄積用動画画像、移動体用動画画像、更にはインターネットなどの準リアルタイムでイン

第1章 序論

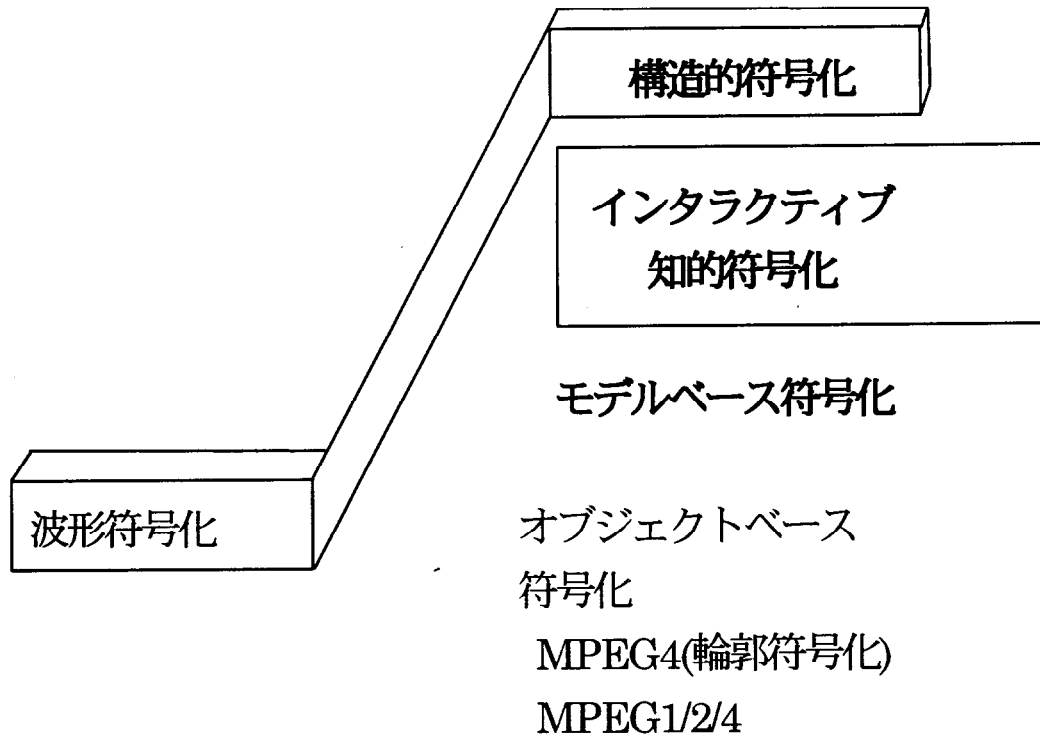
タラクティブな自由形式の画像など様々な形で急速に変遷し、これらをいくつかの情報理論などだけで統一的扱うことには無理があり、個別の方式に対し多くの特化した研究・開発が必要である。

画像の高効率符号化に関する国際標準化活動として、静止画像に対する ISO, IEC の JPEG (Joint Photographic Experts Group)、テレビ電話・会議用動画像に対する CCITT(現、ITU-T)の H.261/H.263、蓄積メディアから通信・放送へも発展した動画像に対する ISO, IEC の MPEG (Moving Picture coding Experts Group)1/2/4 などがある。このような標準化活動においては、理論的な面での検討をベースにした上で、多くの部分で実験的な試行錯誤も加えて検討が進められた。仮に構成されたモデルに対応する計算機シミュレーションによりシステム検討を行い、必要により計算可能な方式をしらみつぶ的に調べ、性能を向上させる観点で最適化を図って来た。処理部分を組み替えていくという改良を行なう手法を組織的に遂行していくことで、改良が蓄積され、実用的で性能の良い標準符号化方式が開発されてきた。

このように、代表的な基本方式に対しては、基礎理論により符号化の限界は与えられるが、個別の画像メディアに対する各種方式については、それぞれの研究を行う必要がある。数ある符号化方式に対し、その一部ずつであっても、理論的に最適化の検討がなされれば、画像の高効率符号化はより最適なものになり、また、理論的な究明がなされるということは、逆に画像信号や個別の画像処理の性質が明確化されるということも意味している。

図 1.1 は画像の高効率符号化研究の発展の模式図である。画像を物理的な信号として、主に統計的な手法に基づき圧縮を行い波形を忠実に再現する、いわゆる波形符号化と呼ばれる方式として発展を遂げてきた。これらは今や国際標準化方式として体系化が進みつつある。しかしながら、画像符号化分野全体に広く目を向ければ、このような波形符号化は、第0ないしは第1世代の符号化として分類するという考えがあり、構造抽出、特徴抽出や分析・合成、認識・合成、知的処理などを行なう第2～第5世代の符号化構想が提案されている。画像符号化はこのように波形符号化から次の世代の構造的符号化へ発展的に研究が進められており、アプリケーションの要請によりいろいろな段階での符号化方式の研究が展開している。画像を単なる物理的な信号として捉えるだけでなく、その構造や意味的な解析に基づく次世代符号化においては、画像情報を単に高効率に圧縮するというひとつの目的に留まるだけではない。即ち、

CG: (Computer Graphics) VRML (Virtual Reality Modeling Language)



動画像: MC+DCT+VLC→H.261/3
(Motion Compensation), Variable Length Coding)

静止画像: KL 変換符号化方式
DCT (離散コサイン変換) →JPEG

2値FAX: MH,MMR→JBIG
(Modified Huffman, Modified Read)

図 1.1 画像の高能率符号化研究の発展経緯

第1章 序論

符号化の前段階において、画像情報を構造的に分析していく過程を介し、対話的な画像の編集・加工処理にきわめて有効な構造情報を得ることが出来る。

顔動画像に特化した分析・合成符号化は従来の波形符号化のように入力した画像信号の圧縮処理のみを行なうという固定概念を破り、人間の認識機能を模し、送受信部に知識を共有させることにより、画期的な高能率化を図ろうとするものである。計算機シミュレーションにより、顔画像の合成がなされ、その潜在的な高性能が次世代符号化方式として評価されてきた。このような分析・合成符号化は既存の技術範囲にとらわれずに、それを超越した飛躍的な位置に新概念—モデルベース符号化ないしは知的符号化—を構築した。モデルベース符号化は、波形符号化のような主に統計的な手法による物理的な信号処理にとらわれるものではない。直接的には一層の高能率化を目指すものであるが、それに付随して画像が構造化されることにより、もたらされる有効な情報は、各種の新しい信号処理、とりわけ対話的な編集・加工などの処理にとって有用な情報を提供出来る。単なる画像の高能率符号化という信号処理から、編集・加工までを含む符号化・復号システムに拡大されたものになっている。分析・合成符号化の実験によりこのような考えも見受けられるが、分析部分では画像の自動解析ないしは認識などを行う部分がいわば飛躍させた部分であり、その研究課題を解決することの必要性を提起する役割を果たした。一方、インターネットなどのマルチメディア情報が多様化したことにより、動画像符号化における従来からある全自動かつ実時間処理の条件は必ずしも必須ではない状況が生まれつつある。非全自動で非実時間処理のインタラクティブ処理を符号化過程に導入して行く新しい概念が必要となって来る。そこには、インタラクティブ操作に関する新たな符号化システム上の問題やインタラクティブ操作そのもののヒューマンインターフェース上の問題が検討されなければならないだろう。

従来の知的符号化においては、全自動処理が未だ達成されていないため、人間がその一部の処理に対して補助を行ない、符号化の過程を完了させるようなことが、事実上必要であった。しかし、従来このような知的符号化では、本来全自動で実時間の符号化を行なうことを目標としており、人間の補助はあるべきでないものとして積極的に取り入れることはなかった。これに対して、インタラクティブ操作は非実時間環境の普及開始に伴い、積極的な意味で取り入れることが重要な課題となって来ている。ここにおいて、インタラクティブ操作というものは人間的であるため、曖昧な処理となり易い性質があり、工学的にどのように数値化ないしは規格化して行くかが重要な問題点として新たにあげられる。

1.2 研究の目的

デジタル画像の高能率符号化における新たな符号化方式の開発や最適化を行なう場合、情報理論等の基礎理論に基づくだけでは十分な結果が得られないことの方が多い。情報理論等に基づく基礎理論は限られており、一方、画像メディアの方はマルチメディア化の中で多様化し、アプリケーションも豊富になっている。本研究では、波形符号化から次世代符号化方式まで多くの研究がなされている背景の中で、発展的な研究のひとつとして次世代符号化方式として高い性能と画像データの構造化の有効性が検証されてきた分析・合成符号化を取り上げ、その実用化へ至る上でのいくつかの課題を検証し、新たにインタラクティブ操作の導入を提案して行く。インタラクティブ操作は従来の知的符号化では、本来あらざる処理としてあまり述べられることがなかった。本研究では、このインタラクティブ操作を逆に積極的に導入することを目指し、知的符号化の実用化上の諸問題の調整を図るとともに、インタラクティブ操作が人間的である事による曖昧性を取り込まないように、インタラクティブな操作量に規範を与え、その規範と各種の誤差の関係を解析し、操作量に規範のあるインタラクティブ知的符号化を提案していく。

符号化対象を顔動画像に限定することにより符号化効率の向上とワイヤフレームモデル等により表現されたパラメータ情報が編集、加工などの付加機能をもたらす方式について検討する。顔動画像への対象の限定は汎用的な符号化では達成できない符号化効率の向上と高機能な信号処理をもたらす。本論文では顔動画像の知的符号化を実用化可能な、非実時間、インタラクティブな環境において再構築し、実用上の諸問題を解決し、ネットワークを用いた伝送実験により検証を行っていく。

本研究内において、従来主に行なわれていた波形符号化を中心とした研究や標準化の過程を見直し、符号化効率の向上を平均二乗誤差を最小化する観点から、KL変換の検討を行い、これに適応信号処理を導入する試みを行った。1枚の静止画像に対し、画素単位に予測符号化を行なうよりも、複数画素の相関を除くブロック符号化方式は効率の向上が期待できる。静止画像に対する波形符号化において、KL変換符号化方式の適応化により、平均2乗誤差最小の観点から最適性を追求し、適応信号処理を導入することにより最適化を行なう。画像の高能率符号化は世界的な規模での標準化が行なわれ、経験論の集積による性能の向上が図られて来た。そのようないわば試行錯誤的な手法に対し、本論文では部分的ではあるが、理論的な基盤と最適化を図る手法を研究し、そのような信頼性の高い要素技術を統合していけば、統一された理論体系

として完成されていくことを目指している。前述したインタラクティブ知的符号化においても、その初期画像を静止画像として符号化するが、静止画像の符号化性能を向上させることも重要な課題となっている。

1.3 インタラクティブ知的符号化方式

符号化効率を向上させるために分析・合成符号化が提案されている。分析・合成符号化は統計的冗長度圧縮を行なう第1世代符号化、動き処理を行なう第2世代符号化等に対し、第3から第5世代の符号化へ展開するものである。顔動画像に特化するという適応的な処理を導入しているとも見做せる。送受信に画像符号化に関する共通の知識を設け、効率的な符号化を行なう。テレビ電話などの実時間通信を想定して始められたものであるが、画像の特徴点などを自動的に抽出することが難解な課題となっている。これを非実時間、インタラクティブな環境で再構築することにより、安定で信頼性の高い実用システムが構成できると考えられる。インタラクティブ操作は従来のこのような課題に対し、単に人間が補助を行うというだけでなく、どのような操作をどの程度行なえば、どのようなパラメータが安定に検出できるのかを明確化しながら導入することが重要である。また、インタラクティブ操作の導入により、単に符号化を行なうだけのシステムから、顔動画像の加工や編集などの付加機能を加えたシステムを構成できるようになる。このようなインタラクティブ操作の概念を導入するため、まず、2次元の顔画像上の特徴点の情報から3次元の動きパラメータを検出するアルゴリズムを開発していく。これは一般的には困難な問題であるが、初期顔画像のワイヤフレームモデルの3次元構造を仮定することや、カメラと被写体の距離や焦点距離の関係から近似を行なえば、その近似は対象としている画像に対しては精度の高い近似になるという状況の仮定に基づいて導出している。また、インタラクティブ操作というものは、オペレータがマニュアルに必要なところだけ操作し、修正を行うのであれば、工学的に規定を与えたことにならない。操作の精度や回数を規定することが必要である。これを明確化するため、操作に付随する誤差を除去するノイズ除去フィルタの開発と精度の検証を行った。更にインタラクティブ符号化ツールの全体を開発し、多数の画像で符号化実験を行いその有効性を検証した。この中で、従来の知的符号化では画像の特徴点や動きパラメータを自動的に抽出することが困難であるため、方式の基本的動作を調べるための符号化・復号の計算機シミュレーションが数多くなされてきた。しかしながら、符号化システムが実用化されることはなかった。知的符

号化システムの多機能化を図ることよりも、基本的な障壁である特徴点の自動検出・追従や動きパラメータの自動検出の解決のための多くの研究がなされるようになった。例えば眼鏡のある顔動画像の符号化、髭のある顔動画像の符号化、非正面画像を初期画像としてワイヤフレームモデルのフィッティングを行なうなどの知的符号化に対する多機能化は課題とはなったものの、積極的に解決を行なう動きはなかった。

本研究で開発するインタラクティブ知的符号化の符号化ツールを用いれば、それをベースに知的符号化の多機能化の研究を開始することが出来る。本研究では、上記知的符号化の符号化ツールの構築後、眼鏡付き顔動画像の符号化を採り上げ、具体的符号化実験や画質と操作量の関係を調べていく。実際に講演などの顔動画像を収録する際には眼鏡をかけた人物がわざわざインタラクティブ知的符号化のために眼鏡をはずして講演を行なうということは考えにくが、そのような状況での眼鏡を電子的に除去する画像処理と画像合成後に再度眼鏡を上書きする処理を実験し、実用的な操作内容で処理を有効に行う基準を導出した。

・適応KL変換符号化方式

画像高能率符号化の符号化性能の極限というものを考える時、いかにそれを理論的に求めるか、又はいかにその極限に向かって実際の符号化方式を開発していくかという課題は興味深いものがある。レート歪み理論やエントロピーは符号化の一種の極限を与える理論的指標であり、方式を分解した基本構成に対する計算は可能だが、近年の複雑化した符号化方式の手続きに対応した理論的限界値を完全な形で算出することは現状では不可能である。例えば、1枚の静止画像に対する符号化方式として前値による予測のみを行い、その予測差分を無歪みにPCM符号化するという最も単純な予測符号化方式に対しては、1次マルコフエントロピーを求めれば符号化の限界が得られることになる。しかし、予測を前値の1画素のみでなく複数画素を用いて適応的に行なったり、予測誤差の適応的な量子化などが組み合わさった形の符号化に対しては、エントロピー計算自体が複雑化し、膨大な組み合わせの状態に対応した画像データを用意することも困難になってくる。また、一般には予測残差には量子化を行い、非可逆な符号化を行なうことが多いため、もはやエントロピーを計算するだけでは不十分で、ビットレートと歪みを同時に考慮する必要がある。更にブロック符号化においては、現実の複雑化した符号化方式の限界値を正確に与えるような詳細な分野をカバーする理論の検討が十分になされていない。このうち、ベクトル量子化に関しては高レート

第1章 序論

においては、スカラー量子化に比較しサンプルあたり最大 0.255 ビットの改善量がありうることを示されている[3]。ただし、高レートであることと、改善量がそれほど大きくないことから、実際の応用は行いにくい。理論的に詳細な手法が検討できない部分は大規模な計算機シミュレーションによる試行錯誤的な手法に基づく検討がなされている。CPU、メモリの技術進歩による低コスト化に支えられ、符号化方式の研究開発は世界的標準化活動へと発展してきた。

KL変換はガウス分布の入力ソースに対し変換後のデータの量子化による原信号と再生信号との誤差を最小化するという意味で最適な変換として知られている。入力共分散行列がその固有ベクトルにより対角化される時、平均2乗誤差が最小になる。単一の変換行列を用いる変換符号化方式に対して、複数の変換行列を用意し、変換の切替えを行なう適応変換符号化方式は、画像の局所的性質に応じて信号のエネルギー分布が更に集中できるため、符号化効率が向上する。KL変換とDCTは類似していると言われるが、本来のKL変換が平均値を分離したベクトルの自己相関である共分散を扱うのに対し、DCTでは平均値分離の手続きを行わない。現実の符号化システムを考えた場合、平均値を分離した純粋のKL変換符号化方式では、変換後の第1番目の係数は0になり、平均値と第2番目以降のデータの量子化を行えばよい。これに対し、KL変換行列を複数個用いて変換の切り替えを行う、適応KL変換符号化方式では、変換後の第1番目の係数は必ずしも0とはならない。符号化システムとしては始めに分離した平均値の量子化と、変換後の係数の量子化が必要になり、サンプル数が1つ多いデータを量子化伝送することになる。そこでこのようなサンプル数の増加を防止するため、平均値の分離を行わない形での適応KL変換符号化方式の構成を行った。

適応KL変換符号化方式と類似した方式に特異値展開符号化方式がある。信号のブロックから特異値ベクトルを求め、そのベクトルの成分の一部を量子化するもので、特異値ベクトルに対応する特異値が急減少する効果を生かそうとしている。これに対し、適応KL変換符号化方式は自己相関行列の固有ベクトルで画像ブロックデータを変換し量子化する構成をとる。適応KL変換符号化方式がエネルギー集中を行なうための最適化を行なっているのに対し、特異値展開符号化方式では特異値ベクトルの量子化を行なうため、エネルギー集中などの2乗誤差を最小化するという観点での最適化を直接的には果していない。また、特異値ベクトルへの展開によってサンプル数が約2倍以上に増加するという点も問題である。そこで、本論文ではこれらを区別し、適応KL変換符号化方式としての検討のみを行なう。符号化対象となる画像は汎用な

ものであるが、その中に存在する局所的な特徴、いわば個性的な部分に適応して符号化を行なうことによって、符号化効率の向上を図ることができる。

1.4 本論文の構成

本論文は、第2章でインタラクティブ知的符号化における初期画像の伝送において必要となる静止画像の波形符号化を取り上げる。まず、KL変換符号化方式の起源と基本構成を整理し、新たに平均値分離を伴わない適応KL変換符号化方式を提案して行く。適応KL変換符号化方式の符号化誤差を適応KL変換符号化方式の変換行列集合の設計における評価関数とし、2乗誤差最小化の観点での最適な適応KL変換符号化方式システムを構築して行く。

第3章では、インタラクティブ知的符号化の基本構成の提案を行う。適応信号処理の究極化の一つとして、顔動画像への特化と第2世代の波形符号化から、第3～5世代の知的符号化への進展において、符号化に伴ってなされる画像の構造化記述情報をアプリケーションに反映していくことを試みる。知的符号化の文献の調査に基づき、符号化と構造化との関係を検討し、第3～5世代の符号化においては、構造化記述を図ることと符号化が強く関連していることを述べる。次に従来の知的符号化の問題点とそれを解決すべく導入したインタラクティブ操作により、新たにインタラクティブ知的符号化システムの構想を提案し、インタラクティブ操作の意義の検討を行い、操作量が曖昧となるという問題点と操作量に規範を与えるという課題を整理する。

第4章では、第3章で提案されたインタラクティブ知的符号化を具体的に構築するため、人物頭部の3次元のグローバル動きパラメータの検出について検討を行なう。まず調査によると、人物頭部の3次元動きパラメータを一般的に求めるアルゴリズムは無く、平行投影による近似を行なうとか、小さい範囲の動きに限定して求めるなどの不完全な手法が開発されているのみであることが分かる。そこで、問題の対象を一般論から人物顔動画像の符号化に限定出来ることを前提条件として積極的に活用することを行い、撮影環境下において、インタラクティブ操作で与えられる情報と合わせ、顔動画像に対する新たな3次元のグローバル動きパラメータ検出のアルゴリズムを開

第1章 序論

発して行く。

また、インタラクティブ操作で発生する入力情報に対するノイズ成分の解析を行い、顔画像の構造を保存しつつ、ノイズのみを選択的に除去していく新たな構造的フィルタの提案と、作用の検討を行う。更に、これらの検討結果を統合し、グローバル動きパラメータ検出におけるインタラクティブ操作に対する精度の評価を行い、操作者の操作に対する精度上の仕様を規格として与えることができるようにした。

第5章では、第3章で提案されたインタラクティブ知的符号化システムのうち局所的な動き乃至は表情などに関わるローカル動きパラメータの検出についての検討を行なう。ローカル動きパラメータを記述する際に、グローバル動きパラメータとの記述順序により、知的符号化の符号化システムが異なって来ることを明らかにする。ローカル動きパラメータの記述において、ワイヤフレームモデル上の頂点の相互関係づけを行い、より少ない指示点で符号化対象人物の特徴を保持する連携規則を構築する。

以上で、インタラクティブ知的符号化システムの構築がなされるが、このインタラクティブ知的符号化ツールを用い、従来なされていなかった知的符号化の多機能化の一つである、眼鏡のある顔動画の符号化実験を行なう。眼鏡画像処理で新たに発生する処理の精度の程度を評価するため、テレビジョンのゴーストと画質評価結果を用いる試みを行っていく。

第6章では以上の成果を具体的に検証する、インタラクティブ知的符号化実験を各種行なって行く。CCITT などの標準的な画像の他に、独自に撮影した諸画像での符号化を行って行く。また、符号化結果をデータベースの一つとして格納し、ネットワーク経由で見ることが出来るような構成を行い、転送実験を行っていく。又、従来の波形符号化の代表的方式である ITU-T の H.263 方式との主観的な比較評価を行い、ビットレートと画質の関係で、提案している知的符号化が H.263 方式と同等乃至は優れた性能であることを確認して行く。

第7章では、人間のインタラクティブな操作量に規範を与えたインタラクティブ知的符号化の提案とその有効性についての全体的なまとめを行う。

第2章 適応 KL 変換符号化方式

2.1 まえがき

画像の構造的な符号化方式であるインタラクティブ知的符号化においても、後の章で述べるように初期画像やパラメータの伝送には従来の波形符号化方式を使用する必要がある。本章では、インタラクティブ知的符号化の研究に先立ち、波形符号化の中で理論的に最適性を有すKL (karhunen Loève) 変換の検討を行なう。

画像の高効率符号化は離散コサイン変換 (Discrete Cosine Transform:以下 DCT と略す) を用いた方式を基本に静止画像、動画像に対する標準化が進み、その応用技術開発が活発化している。DCT 符号化方式は定常な入力ソースに対し最適に近い特性を示している[6]。しかし今後の LSI 技術の進展を展望した場合、複雑ではあるが画像が本来持っている非定常的性質[7]を考慮した新しい符号化方式が有望と予想される。

一方、次世代符号化方式として提案されてきた知的符号化における初期画像の伝送では、波形符号化が用いられているが、DCT 符号化に対し適応信号処理を導入することにより、更に高効率化を図れることが期待できる。

本章ではまず、今までのKL変換符号化方式に関する研究を調査し、理論的な観点で整理を行なう。その際、従来あまり述べられていない特異値展開 (SVD) 符号化方式との差異についても言及する。画像の非定常性に対応し、局所的性質に適応しながらKL変換行列を切替える方式に関しては、従来特定なモデルについての研究がなされて来た。本章では、KL変換符号化方式の特徴である原信号と復号信号との「平均2乗誤差最小化」の評価基準を設計アルゴリズム全体に拡張する。これにより実画像をトレーニングデータとすることができ、画像データ特有の特徴に基づいたKL変換行列を設計する手法を確立した。KL変換行列の形式として、理論的な定義をそのまま用いる入力ベクトルの共分散行列から求まる行列は、適応変換符号化を行なう時に量子化手続きが重複する形になる。これを回避するため、新たに平均値分離を行わない自己相関行列を用いてKL変換行列の導出を行った。これにより効率の向上が得られると共に、従来の単一 DCT 符号化方式との比較が同じ条件で行なうことができるようになる。計算機シミュレーションにより、適応KL変換符号化方式の有効性が確認できた。提案する設計法で求められたKL変換行列から得られる符号化特性はトレーニング画像の違いに依存しない結果が得られ、非定常性に対して汎用なシステムが

第 2 章 適応 KL 変換符号化方式

構成されているものと考えられる。本設計方式の検討において途中の段階で求まる固有値は事実上全て正で求まり、安定な設計が保証できることも示す。

以下、2 節では KL 変換符号化方式の研究の背景の調査を行い、本研究の位置づけを明確化していく。3 節では平均値分離を行なわない形式の KL 変換行列の導出と、それを用いた変換を切替える適応 KL 変換符号化方式を構成する。4 節では、SVD 方式との差異についての検討を行なう。5 節では、適応 KL 変換符号化方式における実画像データを用いた KL 変換行列の設計方式について述べる。6 節で符号化実験と DCT 符号化方式との比較を行う。7 節でまとめを行なう。

2.2 従来の KL 変換符号化研究について

KL 変換はガウス分布の入力ソースに対し、変換データの量子化や切り捨てに伴う平均 2 乗誤差を最小化できる最適変換として知られている[8][9]。直交変換として信号のベクトル g と g^T のテンソル積[10]の形式 $g \cdot g^T$ (以下これを g と g^T のテンソル積と呼ぶ) を対角化する固有ベクトルからなる直交行列 Φ が存在し、

$$\Phi^T \cdot g \cdot g^T \cdot \Phi = \Lambda$$

又は、

$$g^T g \cdot \Phi = \Phi \cdot \Lambda$$

(Λ は対角行列)

となることは線形代数の実対称行列の標準形への変換 (主軸変換) として古くから知られていた[11]。工学への応用ではガウス分布の入力ソースに対して、平均を差し引いた共分散に対する検討がなされ、P.A.Winz[12][13]によると Hotelling, Karhunen と Loève により連続信号の変換 (KL 展開) が発表され、1954 年に A.Koschman が、1960 年に J.L.Brown が別々に 2 乗誤差が最小となることを証明している[11]。離散信号に関しては 1956 年に Kramer と Mathews によって Zonal Sampling の場合の平均 2 乗誤差の最小性が証明されている[12]。但し正変換と逆変換の特定化はまだなされていない。1963 年に J.J.Huang と P.M.Schltheiss により入力が平均 0 のガウス分布の場合に①線形変換の中で逆演算として平均 2 乗誤差最小の観点で最適なものは逆行列である。②正変換として平均 2 乗誤差最小の観点で最適なものは入力の共分散の固有ベクトルである。③固有値の大きさの減少順に量子化ビット配分し、Max の量子化をする時も上の変換が 2 乗誤差最小の観点で最適であることが証明されている[14]。これにより単一の行列を用いる KL 変換符号化の最適性が確立したといえる。その後、入力ソースがガ

第2章 適応 KL 変換符号化方式

ウス分布の場合で Zonal Sampling の時の平均 2 乗誤差最小の観点での最適性の異なる証明は各種発表されている[11]。

変換行列を複数個用意し、画像の局所統計的性質に従って変換の切替えを行なう適応変換符号化方式において、KL 変換を用いる方式は Tasto ら[12][13]によって研究されている。Tasto らはガウス分布の入力画像を分割するモデル化により、単一の変換を行なう場合と複数の変換から選択して行なう場合について Rate-Distortion 関数の上界を求め、ある条件の下では複数変換を選択するインデックス情報を加えても後者のレートが前者のレートを下回ることを導いている。この結果を応用し、実際の画像で符号化シミュレーション等を行なっている。結果として得られた符号化方式では、適応化の分類として直感性に依存した低域、高域、明暗という 3 種の特性に基づいたものを作成したに留まった。Tasto らの実験結果では従来の非適応変換符号化方式に対する優位性が主張されているが、それが適応変換によるものか、適応量子化によるものか明らかにされていない[9][15]。その後、変換行列を切替える方式に関しては Götze[16]らが画像を 8 つの方向成分にモデル化し、適応変換符号化を行なう方式について検討している。この 8 方向適応変換符号化方式においては、解像度チャートの 45 度の斜線成分に対しては DCT 符号化方式より 3 dB 程度優れた性能が得られたが、他の成分や一般の自然画像に対しては DCT 符号化方式と同一の性能は得られていない。Bjøntegaard [17]は 16 種の方向成分を表す自己相関行列から生成した Directional Transform(以下 DT と略す)を用い、方向性が検出されたブロックに DT を適用し、それ以外は DCT 符号化方式を適用する方式を提案している。テレビ会議用の動画の直交変換後のエネルギー集中度で 20%の改善が得られているが、17 種の変換行列の識別を行なうインデックス情報についての記述が明らかでない。また方式としても 16 種類の種類の変換行列の導出の方式が明らかでなく検討の余地がある。Wornell[18]らは平均と分散により特徴となる空間を分類するモデル化を行なっている。得られた各部分集合に対し KL 変換を求めており、モデル化としての進展が見られる。しかし、実画像に対する最適性の関係が密接に図られていない。

本研究では、このような背景の調査に基づき、従来理論的なモデル中心に行なわれてきた適応 KL 変換を実際の符号化方式として具体化するために、平均値分離などの取り扱いを検討するとともに、モデルのみに基づいた適応性の分類ではなく実画像に密接に対応した適応性の分類を行い、また従来 DCT 符号化方式との比較を正確化するため、2次元の可分な変換符号化方式として共通化するなどの検討を行なって行く。

2.3 適応適応可変行列 KL 変換符号化方式

画像を $N \times N$ 画素のサブブロックに分割した行列を $[G]$ とする。縦変換と横変換を行なう $N \times N$ 画素の二つの直交行列を T_C, T_R とする時、

$$F = T_C \cdot G \cdot T_R^T \quad (\text{但し、} T^T \text{ は転置を表す})$$

で定義される直交変換を行った後、量子化、符号化する方式を2次元直交変換符号化と呼ぶ。変換後のデータの状態により異なる量子化ないしは符号化を行なう方式を通常「適応直交変換符号化」方式と呼ぶことが多い。そこで、これと区別するために変換行列 T_C, T_R が画像の局所的性質により適応的に変化する方式を適応的可変行列の直交変換符号化と呼ぶことにする。上記縦変換と横変換はそれぞれ下記のような N 回の1次変換と対応している。すなわち画像行列 $[G]$ の列成分ベクトルまたは行成分ベクトルを g (列ベクトル) としたとき、それぞれ N 個の $T_C \cdot g, T_R \cdot g$ なる行列より成り立っている。直交変換を改めて T とおくと、変換

$$f = T \cdot g$$

を施し、量子化した後、2乗平均誤差を最小とする変換は符号化対象である入力信号 g と g^T のテンソル積[10]の平均

$$E\{g \cdot g^T\}, \quad (\text{但し、}\{ \} \text{ は成分ごとの平均で、以下これを自己相関関数行列と呼ぶ})$$

の N 個の固有ベクトルを行ベクトルの成分とする変換で、KL 変換と呼ばれている [8][9]。ここで、自己相関関数の定義には、平均値を差し引く共分散型[19]と引かない型[20]があるが、本論文では後者に従った。通常、入力画像 $[G]$ の平均値を分離したいわゆる共分散を用いて、符号化の平均2乗誤差の最小性の証明がなされている。ここでは、DCT と同形式で扱うため、平均値を分離しない実画像に対する自己相関行列を用いた。このように平均値を分離しない KL 変換は符号化の過程で次のような特徴を有する。すなわち、 N 次の小ブロックごとに平均値を分離して KL 変換を求めると、第一次の成分は設定した直流成分 (DC) からの偏位分が現れる。従って、変換符号化としての次元は一般には再び N となり、第一次の成分を設定 DC とそこからの偏位残差との二つに分解して送信する形となり、効率が悪くなる。画像信号から、自己相関行列を求め、それから KL 変換を求めると、離散コサイン変換 (DCT) と類似の変換となる。一方、相関係数 $\rho = 0.95$ の定常画像に対する KL 変換は、また DCT と類似しており、ある尺度による性能がほぼ一致している [6]。しかしながら、このような変換行列の求め方は、画像が至るところ定常であるとの前提に基づいたもので、

第2章 適応 KL 変換符号化方式

実際の画像が非定常であることから、画像の性質に正しく対応した変換とはいえない。画像の非定常性に適応した直交変換が適応的可変行列の直交変換符号化方式で、Tastoらによってその基本型が提案され、その効果が検討されてたが、平均2乗誤差最小化の観点で最適化されているとは言えない。

画像の局所的な統計的性質の変化に対応して、ブロックごとに異なる直交変換行列を用いて変換を行なうことにより、効率の向上が期待出来る。図 2.1 は適応的可変行列KL変換符号化方式の一構成ブロック図で、縦横の変換行列はあらかじめ定めたM

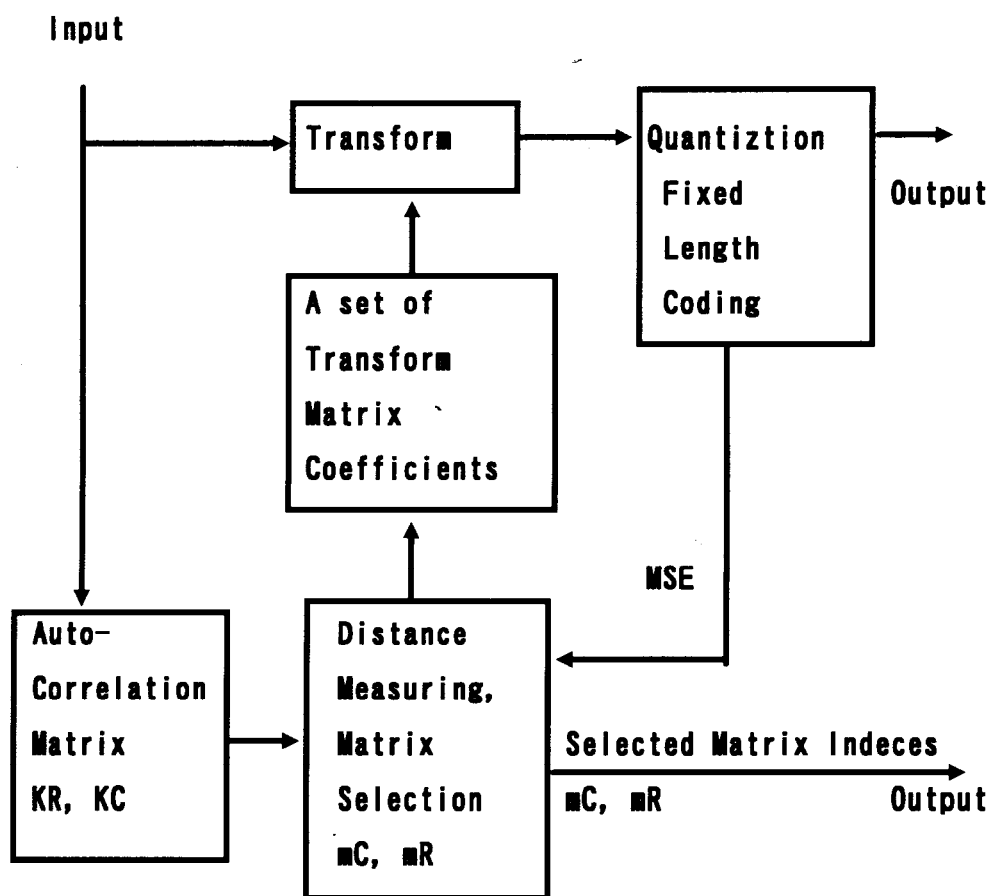


図 2. 1 適応的可変行列KL変換符号化方式

Adaptive matrix-variable KL transform coding

第2章 適応 KL 変換符号化方式

組の変換行列集合の中から選択し、どの変換行列を選択したかを示すインデックス情報をあわせて伝送する。 K_R, K_C は自己相関行列、 m_R, m_C は選択された行列のインデックスを示す。従って、変換符号化の効率向上とインデックス情報の増加のトレードオフを検討する必要がある。また変換行列の設計方法、選択方式、量子化方式、符号化方式、選択した行列を示すインデックスの符号化方式などを検討する必要がある。本論文ではこれらのうち設計法の有効性を確認するために必要なことを検討した後、インデックス情報を含めた上で、変換を切替える適応的可変行列KL変換符号化方式の有効性を調べるために固定の量子化での比較を行なう。

まず、 $N \times N$ 画素のサブブロック行列 $[G]$ 成分を $G(i, j)$, G の行ベクトルを V_{Ri} 、列ベクトルを V_{Cj} ($i, j = 1, 2, \dots, N$) とする。すなわち、

$$V_{Ri} = (G(i, 1), G(i, 2), \dots, G(i, N))$$

$$V_{Cj} = (G(1, j), G(2, j), \dots, G(N, j))^T$$

(但し、 T は転置を表す)

これらの行ベクトル、列ベクトルのテンソル積行列としてそれぞれ、

$$K_{Ri} = V_{Ri}^T \cdot V_{Ri},$$

$$K_{Cj} = V_{Cj} \cdot V_{Cj}^T$$

なるものを定義できる。行ベクトル、列ベクトルのテンソル積行列の成分ごとの平均、

$$K_R = E\{K_{Ri}\}, \quad K_C = E\{K_{Cj}\}$$

を自己相関関数行列とし、その固有ベクトルを列ベクトル

$$\phi_{Rk}, \phi_{Ck} \quad (k = 1, \dots, N)$$

として求める事が出来る。対応する固有値 (後述する定理1により固有値は事実上全て正になる)

$$\lambda_{Rk}, \lambda_{Ck} \quad (k = 1, \dots, N)$$

が大きい順に並ぶよう $\lambda_{Rk}, \lambda_{Ck}$ を整列させておく。このとき、

$$\phi_R = \left[\begin{array}{c|c|c|c|c} \phi_{R1} & & & & \\ \hline & \phi_{R1} & & & \\ \hline & & \dots & \dots & \\ \hline & & & & \phi_{RN} \\ \hline \end{array} \right]^T$$

$$\phi_C = \begin{bmatrix} | & | & | & \dots & | \\ \phi_{C1} & \phi_{C1} & \dots & \dots & \phi_{CN} \\ | & | & | & \dots & | \end{bmatrix}^T$$

がそれぞれ横変換、縦変換の行列となり、可分 2 次元 DCT と行列乗算が同形式の KL 変換

$$Y = \Phi_C \cdot X \cdot \Phi_R^T$$

が定義できた。また、KL 変換は従来平均値分離型の自己相関関数行列で構成されてきたが、この構成は平均値分離を行わない構成がとれるため、平均値伝送のオーバーヘッドがなくなり、現在最も広く使用されている DCT 方式との比較が行列式の変更だけで可能となるという特徴を有する。ここで、上記で定義される KL 変換が種々の画像に対して事実上 N 個まで求まることを保証する次の定理を述べておく。

[定理 1]

画像から作る自己相関関数行列 K_R, K_C は、それぞれ少なくとも N 個 1 組の N 元の 1 次独立なベクトルが画像ブロックに含まれて入れば、他に 1 次従属なベクトルからなる画像ブロックから求まるテンソル積 $g \cdot g^T$ を加算して平均しても、階数は N となり固有値 $\lambda_{RK}, \lambda_{CK}$ は全て正となる。

証明：

テンソル積形式の行列

$$K_{Ri} = V_{Ri}^T \cdot V_{Ri}$$

$$K_{Cj} = V_{Cj} \cdot V_{Cj}^T$$

(V_{Ri} は行ベクトル、 V_{Cj} は列ベクトル)

は対称行列であり、その各 N 個の平均

$$K_R = E \{ K_{Ri} \}, \quad K_C = E \{ K_{Cj} \}$$

も対称行列であるので、みな半正定値である。すなわち、任意の列ベクトル x に対し、

$$\begin{aligned} x^T \cdot K_{Ri} \cdot x &= x^T \cdot V_{Ri}^T \cdot V_{Ri} \cdot x \\ &= (V_{Ri} \cdot x)^2 \geq 0 \end{aligned}$$

$$\begin{aligned} x^T \cdot K_{Cj} \cdot x &= x^T \cdot V_{Cj} \cdot V_{Cj}^T \cdot x \\ &= (V_{Cj}^T \cdot x)^2 \geq 0 \end{aligned} \tag{2.1}$$

である。したがって、

第 2 章 適応 KL 変換符号化方式

$$x^T K_R \cdot x \geq 0$$

$$x^T K_C \cdot x \geq 0$$

である。ここで、正と負の両方の固有値が存在する時などでは半正定値であって、かつ正定値でも 0 でもない不定符号の場合があることに注意する[21]。

以下、正則行列の列ベクトルを取り出して作ったテンソル積を平均した行列が正值になることを検証する。

もし、あるベクトル $x_0 \neq 0$ が存在して、 $x_0^T \cdot K_R \cdot x_0 = 0$ となるならば、

$$x_0^T \cdot K_R \cdot x_0 = x_0^T \cdot (K_{R1} + \dots + K_{RN}) \cdot x_0 / N = 0$$

であるが、式 (2.1) より各

$$x_0^T \cdot K_{Ri} \cdot x_0 = (V_{Ri} \cdot x_0)^2 = 0$$

となる。従って、

$$V_{Ri} \cdot x_0 = 0, \quad \text{for } i=1, \dots, N$$

となるが、成分では

$$V_{R11} \cdot x_{01} + V_{R12} \cdot x_{02} + \dots + V_{R1N} \cdot x_{0N} = 0$$

$$V_{R21} \cdot x_{01} + V_{R22} \cdot x_{02} + \dots + V_{R2N} \cdot x_{0N} = 0$$

.....

$$V_{RN1} \cdot x_{01} + V_{RN2} \cdot x_{02} + \dots + V_{RNN} \cdot x_{0N} = 0$$

従って、

$$\tilde{V}_{R1} = \begin{bmatrix} V_{R11} \\ V_{R21} \\ \dots \\ \dots \\ V_{RN1} \end{bmatrix}, \dots, \tilde{V}_{RN} = \begin{bmatrix} V_{R1N} \\ V_{R2N} \\ \dots \\ \dots \\ V_{RNN} \end{bmatrix}$$

とすると、

$$x_{01} \cdot \tilde{V}_{R1} + x_{02} \cdot \tilde{V}_{R2} + \dots + x_{0N} \cdot \tilde{V}_{RN} = 0$$

でこれは、画像の N 個の行ベクトル V_{Ri} からなる行列が正則でないことを意味している。

一方、もし画像の N 個の行ベクトル V_{Ri} が一次独立なら、いかなるベクトル $x_0 \neq 0$ に対しても、

$$x_0^T \cdot K_R \cdot x_0 = 0$$

とはならず、従って、

$$x_0^T \cdot K_R \cdot x_0 > 0$$

となる。つまり、N X N の画像行列が正則ならその行ベクトルから作られたテンソル

第 2 章 適応 KL 変換符号化方式

積行列は正定値であり、従って固有値は全て正となる。

各 $N \times N$ 画素ブロックごとのテンソル積行列を更に複数画面ないしは全画面平均した場合、式 (2.1) の非負性から同様のやり方で、少なくとも N 個の 1 次独立の行ベクトルと列ベクトルが含まれていれば、そこから作られる自己相関関数行列は階数が N で固有値も全部正となることが保証される事が分かる。(証明終)

この定理の意味は自己相関関数行列を求める時の平均の演算において、平均の対象となる個々の画像ベクトルの中に N 元の 1 次独立な N 個のベクトルが最低 1 組あれば安定な解を求めるのに充分であることであり、他に全白、全黒、市松模様など局所的にはテンソル積行列が正則でないブロックが多く存在しうるが、それらがいくら多く存在しても悪影響を受けないということである。証明のやり方から、1 次独立な N 個のベクトルは特定の 1 画像ブロック内に存在する必要なく、平均をとる画像全体で最低 1 組あれば良い事が分かる。これは本方式の特徴の 1 つで、設計において N 個の解が保証されているため、特殊処理を必要としない扱いが可能である。2 次元可分変換としないで N^2 の 1 次元 KL 変換を用いる方式や SVD などでは階数が減少する恐れがあることが報告されている[8][22][23]。

2.4 SVD 符号化方式との差異の検討

KL 変換と同様に固有値計算を用いる符号化方式に SVD 符号化方式がある[23]。入力信号のベクトル g の離散自己相関 $R = g^T g$ は固有ベクトル Φ を用いた主軸変換により対角化でき、 $\Phi^T \cdot R \cdot \Phi = \Lambda$ 、ないしは $R \cdot \Phi = \Phi \cdot \Lambda$ となる。変換行列の形を予め取り決めておくか相対的に少ない情報で伝送しておき、 g を Φ で変換した $\Phi \cdot g$ を主たる情報として符号化・伝送するのか適応 KL 変換符号化方式である。 Φ と Λ 又はそれらを一体化した $\Phi \cdot \Lambda^{1/2}$ 等のみを符号化・伝送するのが SVD 符号化方式である[23]。

SVD 符号化と適応 KL 変換符号化方式は主たる伝送すべき対象が、基底か基底で変換された変換面のデータかの違いを有する。KL 変換されたデータは無相関でエネルギー集中度が最も高く、従って平均二乗誤差最小の観点で最適性が証明できるが、変換基底そのものは相関もあり、従ってエネルギー集中度も高くない。この変換基底

第2章 適応 KL 変換符号化方式

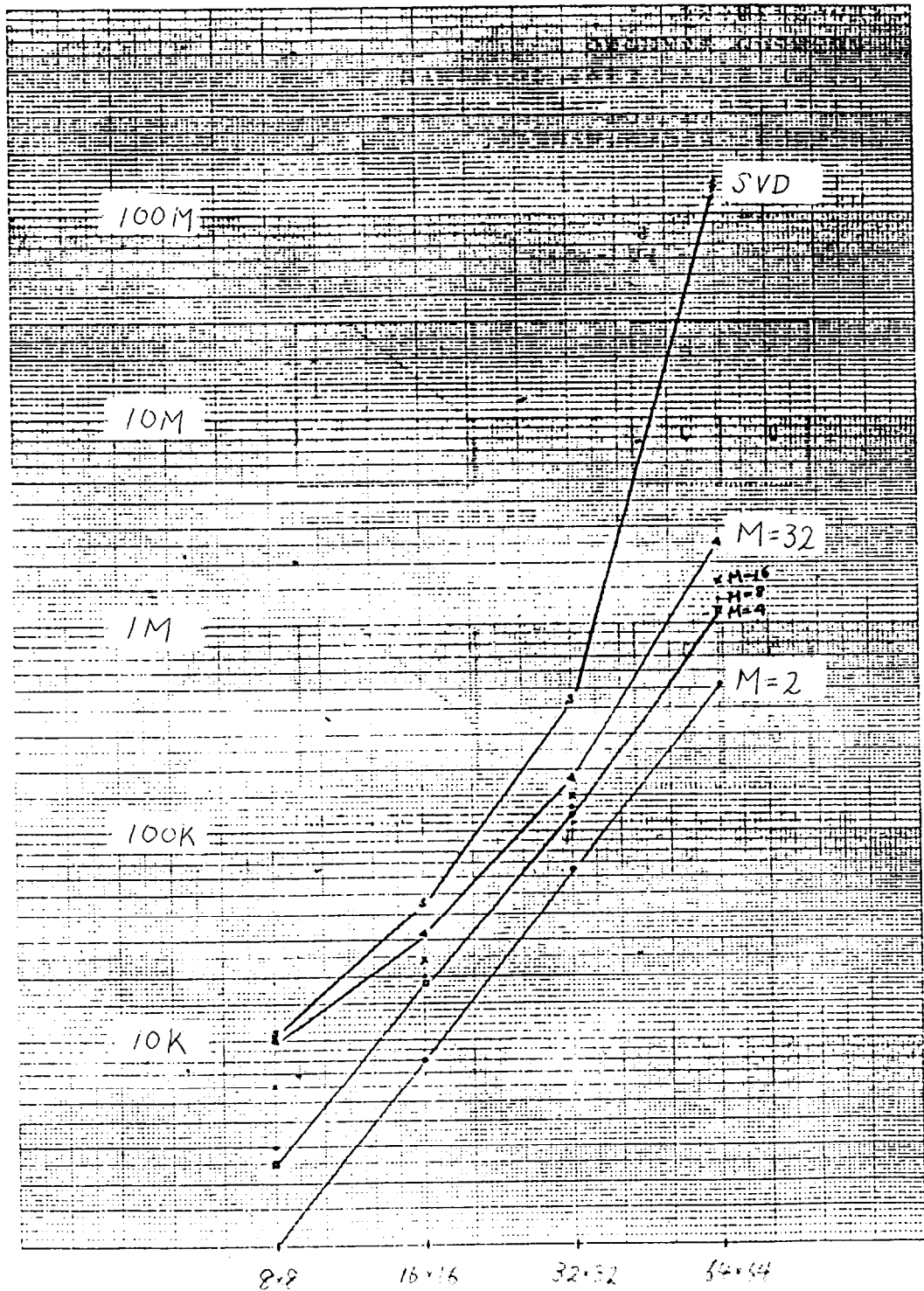


図 2. 2 適応 KL 変換符号化方式と SVD 符号化方式の演算回数の比較

Comparison of operation number between adaptive KL transform coding and SVD coding

第 2 章 適応 KL 変換符号化方式

は対応する固有値が高次になるほど減少するという点は圧縮との関連性はあるが、例えば DCT 変換基底行列のデータは行方向について相関があり、SVD 係数もこれと類似した相関を有することになる。変換符号化では変換基底の伝送量に比べ変換後のデータの伝送量の方が多いのに対し、変換後のデータ伝送を 0 とし、変換基底の伝送のみを行なうのが SVD 符号化方式である。原画像、変換基底、変換後データの中で圧縮に適した信号は「変換後データ」であり、「原画像」や「変換基底」は圧縮には適さない。「変換後データ」の割合の多い変換符号化の方が効率的になるのは明らかである。ここに、SVD 符号化方式の根本的な問題が存在する。

SVD 符号化方式の第 2 の問題点は実行時の演算量の多さである。SVD 符号化方式と適応 KL 変換符号化方式の基本構成は符号化時に固有ベクトルを算出するか、設計において固有ベクトルを算出するかの点で異なっている。両符号化の演算量を乗算回数により評価した結果を文献[12]より下の図 2.2 に引用する。以下の節で詳述する適応 KL 変換符号化方式に対し、SVD 符号化方式は各パラメータにおいて、数倍から数十倍の演算を必要としている。実際、圧縮を目指した試みはなされているが、効率的な結果は得られていない[9][17][22][24]。

適応 KL 変換符号化方式と SVD 符号化方式の違いを表 2. 1 にまとめた。 $N \times N$

表 2. 1 適応 KL 変換符号化方式と SVD 符号化方式との比較

	適応 KL 変換符号化方式	SVD 符号化方式
送出力	直交変換後の変換データ index 情報	信号を特異値展開した 特異値ベクトル (固有ベクトル) 特異値 (固有値)
相関	直交変換という座標軸の 回転により無相関化されている	特異値ベクトルの成分間には 相関がある
集中度	符号化 2 乗誤差が最小となるとい う意味で最適である (エネルギー集中度が最も高い)	エネルギー集中度について最 適性は無い
サンプル 数の変化	なし： $N \times N \rightarrow$ $N \times N + \text{index 情報}$ (変換は直交変換でサンプル数 の増加は無い)	増加： $N \times N \rightarrow N \times N \times 2 + N$ (2 倍以上の空間への非線形 な拡大写像である)

画素の一つのブロックの符号化を行なう時、SVD符号化方式ではSVD展開により $N \times N \times 2 + N$ に源情報が拡大している。SVD符号化方式では圧縮の目的に反してサンプル数を増加させるという非線形の前処理をおこなっている。この冗長度付加の処理は相関のある2つの $N \times N$ 点の空間への写像であり、符号化の対象ととして適切とは言えない。2次元のSVD符号化方式では実際は N 個の特異値と $N \times N \times 2$ の特異ベクトル空間の約半分を伝送すれば、残りの特異値は計算により算出可能であり、また特異値は最大値から急減少しているので、低レートの非可逆符号化は可能である。しかしながら、表 2.1 にあるように、符号化対象は変換符号化でいえば変換係数行列に相当するものであるため、相関が強く、効率的な符号化は行なう事が出来ない。

2. 5 変換行列の設計

KL変換が平均2乗誤差を最小化させるという観点で最適であることを保ちながら、適応的行列可変の方式を構築するための検討を行なう。Tasto らはレート歪み曲線の上界（下界ではないので必ずしも最適性と結びついているわけではない）を実現する分割を与えているが[13]、ここでは平均2乗誤差の極小化を目指す。画像の非定常性に適合した形で解析し、変換行列の設計を行なう方式をいくつか検討することにする[25]。

2. 5. 1 自己相関関数行列の距離による設計

自己相関関数行列の距離に基づいたKL変換行列の設計方式は、変換行列の設計を自己相関関数行列空間内で行なう方式である。図 2.3 に設計のフローチャートを示す。設計のために用意したトレーニング画像から抽出された $N \times N$ 画素の画像ブロックごとの行と列のベクトル V_{Ri} , V_{Cj} に対し、テンソル積行列 K_{Ri} , K_{Cj} を求め、ブロック内で対応する成分ごとに平均し、自己相関関数行列 (Autocorrelation) の集合、 $K_R(IT)$, $K_C(IT)$ を得る。但し、 IT はトレーニング画像の $N \times N$ 画素ブロックの一つを表す。次に上で求めた自己相関関数行列の集合を、ある距離 (distance) を評価関数としてそれぞれ M 個の部分空間に分割される。まず、 $K_R(IT)$, $K_C(IT)$ の集合を初

第 2 章 適応 KL 変換符号化方式

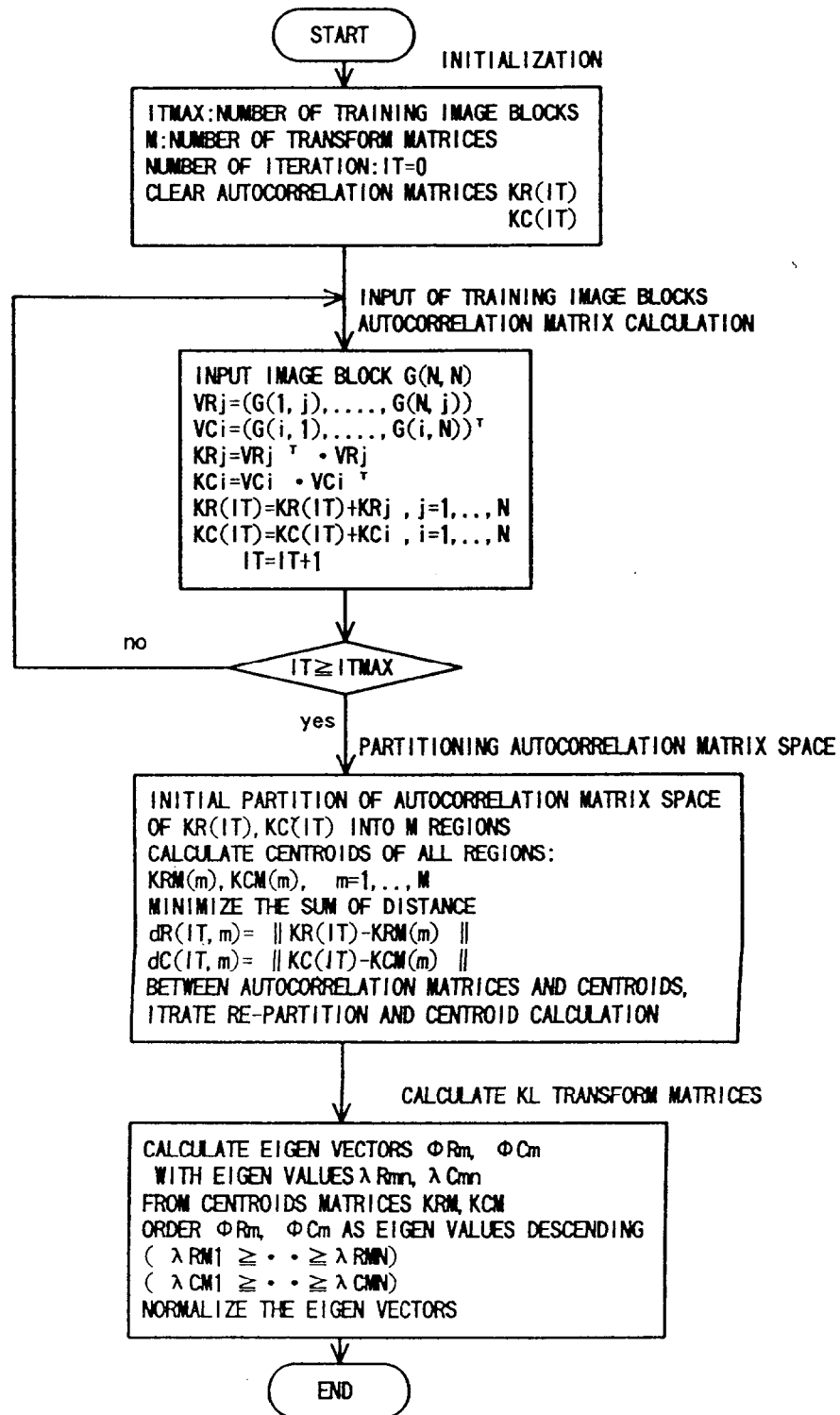


図 2. 3 KL 変換行列集合設計の基本フローチャート
(自己相関関数行列空間での距離を評価関数とする設計)

Flowchart of basic KL transformation matrix design

(Using distance function for auto-correlation matrix space)

期分割し、各分割領域の重心（成分ごとの平均）

$$K_{RM}(m), K_{CM}(m), m = 1, \dots, M$$

を求める。以下、添字 m は分割された領域の一つを表現している。行列の対応する要素ごとの 2 乗和の平均の平方根で定義される距離、

$$d_{R1}(IT, m) = \left[\frac{1}{N \cdot N} \sum_{p,q} \{K_R(IT)(p,q) - K_{RM}(m)(p,q)\}^2 \right]^{1/2} \quad (2.2)$$

を評価関数として再分割を行なう。これは各行列を N^2 次元のベクトルと見て、自己相関関数行列の集合を距離 d_{R1} を使って LBG アルゴリズム [26] で M 個に分類することに相当する。 K_C についても同様にして距離

$$d_{C1}(IT, m)$$

が定義できる。距離 $d_{R1}(IT, m), d_{C1}(IT, m)$ は再分割と共に減少するので、再分割の前後を添字 old, new で表すと、その差分

$$\delta_1 = d_{R1,old} - d_{R1,new}$$

$$\delta_2 = d_{C1,old} - d_{C1,new}$$

が正のある値 ε 未満になった極小点で収束とみなす。

このようにして、 K_R, K_C の各 M 個の領域とその重心 (centroid) である代表自己相関関数行列 K_{RM}, K_{CM} が自己相関関数行列空間での平均 2 乗誤差極小の基準で定めることができた。次に KRM, KCM に対する固有列ベクトルを成分としてもつ行列

Φ_{RM}, Φ_{CM} と固有値

$$\lambda_{Rmn}, \lambda_{Cmn}, m=1, \dots, M, n=1, \dots, N$$

を求める。このとき、固有値は大きい方から順に、

$$\lambda_{Rm1} \geq \dots \geq \lambda_{RmN} > 0 \quad (2.3)$$

$$\lambda_{Cm1} \geq \dots \geq \lambda_{CmN} > 0 \quad (2.3b)$$

と整列されており、固有ベクトル行列の成分ベクトルも対応する固有値の順に整列されているものとする。

ここで、定理 1 より上記代表自己相関関数行列はトレーニング画像の数と分割数 M を適当にとれば、自然画像中の 1 次独立なベクトルを含むのに十分に多くの平均がとられることになるため、式 (2.3), (2.3b) のように固有値は全て正となり、行と列に対しそれぞれ相異なる固有ベクトルを N 個ずつ求めることができる。

この方式は自己相関関数行列の空間内に閉じた設計法であるため、演算量が少ないという特徴がある。また、横変換と縦変換の設計時の反復を同時に並行して行なうことができる。平均 2 乗誤差最小の意味で最適ではないが、自己相関関数行列の距離は

変換誤差の距離と連続関数の関係にあり、近似解としての存在意義はある。

2.5.2 変換誤差評価基準による設計

(可分量子化での設計)

変換行列の設計において、Zonal Sampling した時の現実の誤差を評価基準とすることにより、KL 変換の平均 2 乗誤差最小の特徴と呼応した統一的な設計が可能となる。図 2.4 に量子化を可分とした状態で行変換と列変換を独立に設計する変換誤差評価を用いた設計手順を示す。

初期分割を行い、それぞれ M 個の重心行列 $K_{RM}(m)$, K_{CM} を求める。この分割に対し最適変換は KL 変換 Φ_{RM} , Φ_{CM} であり、これを求める。各自己相関関数行列と代表自己相関関数行列との距離を、

$$d_{R2}(IT, m) = \text{MSE}_i \left\{ \Phi_{Rm} \cdot V_{Ri}(IT)^T - Q \left[\Phi_{Rm} \cdot V_{Ri}(IT)^T \right] \right\} \quad (2.4)$$

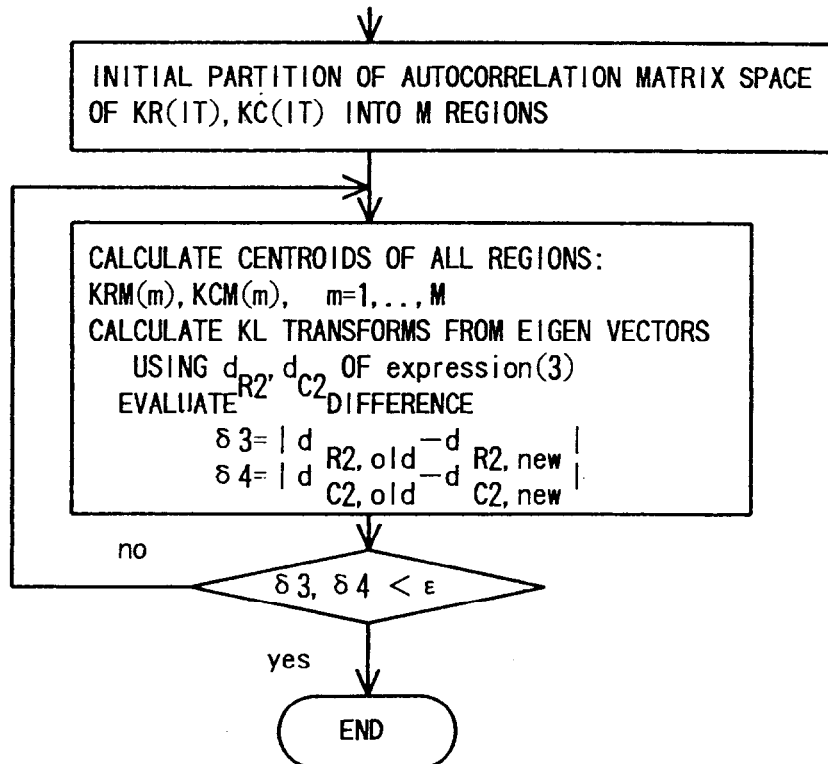


図 2.4 自己相関関数行列空間の分割

(変換誤差を評価関数、非可分量子化での設計)

第 2 章 適応 KL 変換符号化方式

と定義する。但し、MSE はベクトルの成分ごとの 2 乗和を計算するもので、更に添字 i に対応し各ブロック内の N 個のベクトルにわたり平均をとるものとする。Q は一般に量子化関数で、ここでは特に Zonal Sampling する場合に対応して、変換データのうち、初めから k 個採用する場合を仮定している。すると、

$$\begin{aligned}
 d_{R2}(IT, m) &= \frac{1}{N} \sum_{i=1}^N \sum_{l=k+1}^N \left\{ \phi_{Rml}^T \cdot V_{Ri}(IT)^T \right\}^2 \\
 &= \frac{1}{N} \sum_{i=1}^N \sum_{l=k+1}^N \left\{ \phi_{Rml}^T \cdot V_{Ri}(IT)^T \cdot V_{Ri}(IT) \cdot \phi_{Rml} \right\} \\
 &\quad (\text{文献[27]による}) \\
 &= \frac{1}{N} \sum_{i=1}^N \sum_{l=k+1}^N \left\{ \phi_{Rml}^T \cdot K_{Ri}(IT) \cdot \phi_{Rml} \right\} \\
 &= \sum_{l=k+1}^N \left\{ \phi_{Rml}^T \cdot K_R(IT) \cdot \phi_{Rml} \right\} \tag{2.5}
 \end{aligned}$$

となる。また、最終式の固有列ベクトル ϕ_{Rml}^T, ϕ_{Rml} 等は同式内の自己相関関数行列 $K_R(IT)$ とは

同一分割内にないため、固有値にまでには、まとまらない。

各自己相関関数行列と代表点との距離が m に関して最小となる代表点を所属させるという再分割を行なうと、系の 2 乗誤差は非負で 1 つ前の分割と同じか小さい。従ってこの操作を反復的に行なうと、必ず収束する。この手順により、トレーニング画像と分割数を与えた時、平均 2 乗誤差を最小化する設計方式が構築できた。この設計方式は行と列を独立に設計しており、Zonal Sampling のマスクは図 2.5(a) のような可分変換に対応した長方形である。

変換はもともと可分の仮定であったが、量子化の段階で図 2.5(b) のような非可分変換に対応したな斜め領域の量子化を設計に反映するためには、行と列の従属性を考慮する必要がある。以下ではこの従属性を付加下場合についての検討を行なう。

2.5.3 変換誤差評価基準による設計

(非可分量子化での設計)

図 2. 6 に非可分変換に対応した量子化により、行と列の変換の設計に従属性が出る場合の設計手順を示す。2.5.2 との大きい違いは距離関数の定義で、その時点で求まっている最新の KL 変換を用い、対象ブロック IT の縦、横の自己相関関数行列に対し、対応する画像データ G を用い行と列の全組み合わせからなる距離関数

$$d_{R3}(IT, m) = \underset{p, q}{MSE} \left\{ \Phi_{Cm1} \cdot G(IT)(p, q) \Phi_{Rm2}^T - Q \left[\Phi_{Cm1} \cdot G(IT)(p, q) \cdot \Phi_{Rm2}^T \right] \right\} \quad (2.6)$$

から最小の距離を与える領域 m_1, m_2 に所属させる再分割を行なう。

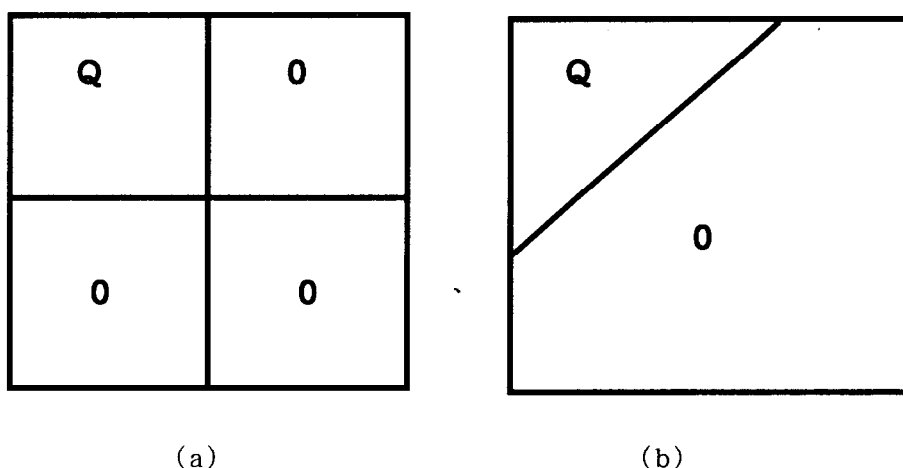


図 2.5 ゾーナルサンプリング領域
Zonal sampling regions.

2.6 計算機による実験

計算機シミュレーションにより、符号化特性を求め、非適応な DCT 符号化との比較を行った。ここでは上で提案した行列を切替える適応方式の有効性を確認するための実験を行なう。従って、適応量子化や可変長符号化の効果を除いた基本比較を行なう。適応量子化や可変長符号化を加えれば性能は向上するが、まず、適応変換のみの

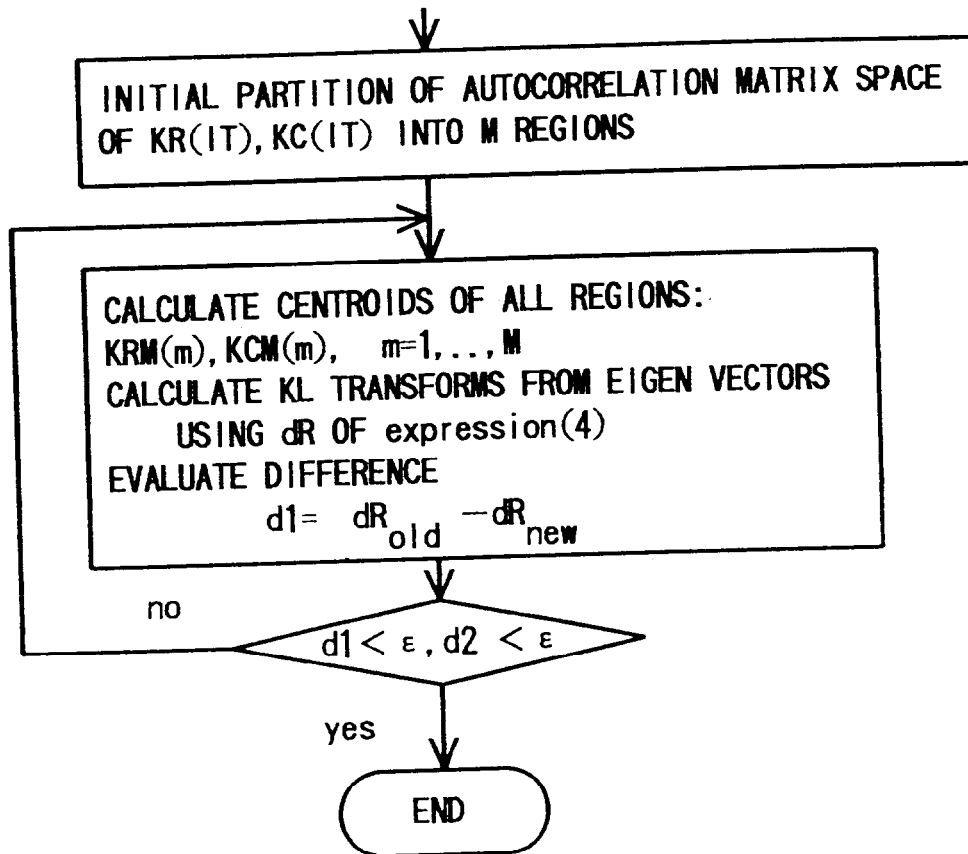


図 2.6 自己相関関数行列空間の分割
(変換誤差を評価関数、非可分量子化での設計)
Partitioning auto-correlation matrix space
(Evaluating transform error for non-separable quantization).

第2章 適応 KL 変換符号化方式

表2.2 符号化性能比較 (S/N、1 bpp、単位dB)

方式 入力	離散 コサイン 変換 符号化	適応的可変行列 KL 変換符号化 トレーニング画像		
		Girl	Couple	Aerial
		Girl	31.6	33.7
Couple	31.5	33.8	34.0	33.5
Aerial	28.4	29.7	29.7	29.7

表2.3 kの値に関する性能比較 (S/N、単位dB)

k bit/pel	3	4	5	初期分割
1.5	36.456	36.494	36.483	36.080
1.0	33.745	33.759	33.745	33.617
0.75	31.495	31.488	31.472	31.441

N=8, M=16, モノクロGirl

表2.4 量子化ビット配分(原形)

	8.1	7.5	6.3	5.2	4.1	3.1	2.1
8.11	7.8	6.5	5.4	4.2	3.2	2.2	1.75
7.4	6.4	5.5	4.4	3.3	2.3	1.72	1.5
6.3	5.4	4.3	3.5	2.4	1.73	1.5	1.25
5.2	4.2	3.4	2.4	1.73	1.5	1.25	1.0
4.1	3.3	2.3	1.73	1.5	1.25	1.0	1.0
3.2	2.2	1.72	1.5	1.26	1.0	1.0	1.0
2.1	1.71	1.5	1.25	1.0	1.0	1.0	1.0



図 2.7(a) 復号画像 (DCT 符号化, 1ビット/画素))

Decoded image (DCT coding, 1bit/pel)



図 2.7(b) 復号画像(適応KL変換符号化方式,1ビット/画素)
Decoded image (Adaptive KL transform coding, 1bit/pel)

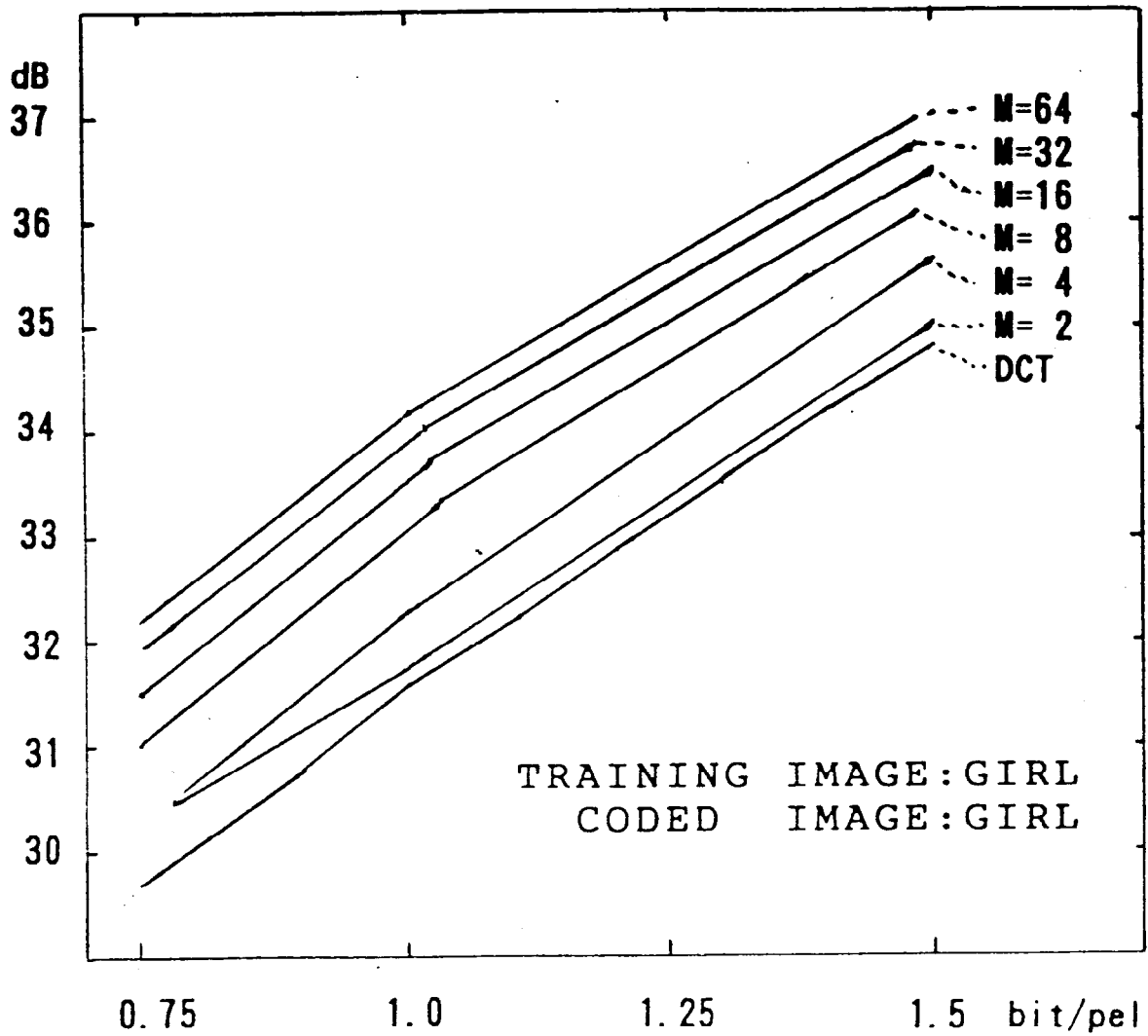


図 2.8 変換行列の数Mに関する S/N 特性(N=8, k=3)

S/N performance for M(N=9, k=3)

第2章 適応 KL 変換符号化方式

表2.5 量子化ビット配分用乗数

bit/pel	DCT	適応KL変換符号化方式
1.5	0.62416810	0.58220285
1.0	0.47619045	0.43478262
0.75	0.37324136	0.31362385

効果を検証することがまず、重要である。また、適応量子化や可変長符号化と一体化した形で比較を行なうと、行列を切替えるという適応変換効果が不明確になるという問題もある。適応的可変行列KL変換は主に2.5.2の設計方式を用い性能を調べた。設計と符号化はCrayのスーパーコンピュータでCPUタイムで約30分要した。このうち設計の方に大半の時間が費やされている。表2.2、表2.3、図2.7、図2.8に結果を示す。量子化は固定で正負対称ミッドライザ型³の線形量子化とし、第一成分は視察による固定ビットとした。他の成分はDCT変換後の係数の分散分布の対数より近似したビット配分を原形とし、これに定数を乗じて切り捨てた結果を各レートのビット配分とした。表2.4に量子化ビット配分の原形を、表2.5に反復的に求めた乗数を示す。DCTは適応的可変行列KL変換の縦、横の変換行列が共にDCT行列の場合に相当し、他の量子化条件は同一とする。すなわち1つのビットレートに対し量子化手法はDCTに1種、適応的可変行列KL変換に1種である。適応的可変行列KL変換では、 $M \times M$ 個の全組み合わせ変換に対し、固定長の量子化を行った時の量子化誤差最小の行と列の組を選択することとし、どの行列を選んだかを示すインデックス情報を付加し、そのビット数だけ符号化ビット数を削減している。インデックス情報は固定長符号化とし、

$$\left(\log_2 M \right) \times 2 \text{ビット/ブロック}$$

となる。

Sidbaの原画はカラーRGBの各色がPCM6ビット精度であるが、RGB画像から実数の色度変換式により求めた8ビットの輝度信号を用いた。表2.2は1ビット/画素においてDCTとの比較、設計法2.5.1、2.5.2、2.5.3の比較、設計法2.5.2で

第 2 章 適応 KL 変換符号化方式

設計に用いるトレーニング画像による性能比較についてまとめている。3 画像ともトレーニング画像と入力一致したときの結果が最も良いが、一致しない時も大きい性能劣化はない。

表 2.3 に設計法 2.5.2 において、初期分割で設計を停止した場合の性能、評価関数式 (2.5) の反復ごとの減少比率が充分小さくなるまで収束させる時の k の値による性能の比較を示す。Zonal Sampling の範囲を示す k の値はビットレートと関連があるが、縦変換と横変換を別々に 1 次元で設計し、実際の量子化は 2 次元で行なうことによる違いが生じていると思われる。DCT との比較ではスプリット法による初期分割を行なって適応的可変行列の形にしたことにより既に性能向上の大部分が達成されており、その後極小値まで収束させることによる効果は小さい。表 2.2 の設計 2.5.3 の効果も同様に小さい。

トレーニング画像が COUPLE の場合の画像 GIRL の復号画像の例を図 2.7 に示す。

次に、図 2.8 に縦、横の各変換行列の数 M に関する性能比較を示す。DCT に対し $M=16$ で約 2 dB、 $M=64$ で約 2.5 dB の性能向上が得られている。固定の量子化による基本性能比較であるため、国際標準方式等に比べ性能は悪いが、変換行列を適応化することにより、インデックス情報を付加した公平な比較において約 2 dB 以上性能が向上することが示され、適応的に行列を可変とする方式の有効性が確かめられた[28]。

2.7 むすび

KL 変換符号化方式を平均 2 乗誤差最小化という意味で最適である。数学的には 2 次形式の固有値問題として扱われ、工学的には連続な自己相関関数についての証明と離散された自己相関関数行列についての証明等がある。これらの系列を整理し、新しい適応的可変行列 KL 変換符号化方式の提案を行った。画像の非定常性を考慮した適応的可変行列 KL 変換符号化方式について、KL 変換の持つ平均 2 乗誤差最小の観点での最適性を踏襲した設計を開発し、誤差の評価関数についての検討を行った。

この中で、本方式によれば設計においても固有値は事実上常に全数正の固有値が求まり、安定な設計を保證できることも示した。また、従来から使用されている特異値展開符号化方式と適応 KL 変換符号化方式の差異について検討を行い、適応 KL 変換符号化方式の方が平均 2 乗誤差最小化の観点に合致していることを示した。

実験では非適応な単一 DCT 変換符号化と条件をそろえて比較し、KL 変換を適応的

第 2 章 適応 KL 変換符号化方式

に切替えることの有効性を確認した。設計に用いるトレーニング画像による影響は小さく、変換行列の数が増加すると性能が向上することが分かった。変換行列の数 $M=16$ で単一 DCT に比べ約 2 dB の S/N の向上が、 $M=64$ では約 2.5 dB の向上が得られた。

今後は、可変長符号化を導入した最適設計、動画像に対する最適設計等を行なう必要がある。

第3章 インタラクティブ知的符号化の基本構成

3.1 まえがき

知的符号化は、低ビットレートにおいて高画質を実現することが可能な次世代符号化方式として研究されて来た。ワイヤフレームモデルのフィッティング、動きの検出、特徴点の追従等、画像の自動認識技術における問題点を解決することが、課題としてあげられている。動画像知的符号化を、テレビ電話・テレビ会議などに応用する場合に、画像の実時間自動認識は必須条件といえる。一方、蓄積系メディア、電子メール、WWW等への応用では、非実時間での処理が許容され、全てを自動処理する必要はない。すなわち、現状の画像自動認識技術の性能が十分でないことによって課せられる信号処理内容に関する負荷条件を除去することができる。このように負荷を低減することにより、知的符号化を早期に実現できるばかりか、新しい信号処理技術の研究・開発が可能となって来る。筆者らは、このような観点から顔動画像のインタラクティブ知的符号化方式について検討を進めてきた。インタラクティブ符号化の基本提案を行い[29]、グローバル動き検出、ローカル動き検出、眼鏡のある顔画像が入力された時に必要となる眼鏡除去のアルゴリズムなどをについての研究とそれに起因する各種の課題の解決を行なってきた。本章では、知的符号化からインタラクティブ知的符号化へ至る過程においてなされた検討と構想の中から基盤となる技術を取りあげて述べることにする。インタラクティブ符号化の基本構成について整理するため、まず画像の構造化と高能率符号化の関係についての検討を行い、構造化の意義と重要性を明らかにする。次に、知的符号化の従来の研究や動向を調査し、本研究の位置づけを明らかにする。それらを基に、インタラクティブ知的符号化の基本構成の提案を行い、従来の知的符号化とインタラクティブ知的符号化の違いやインタラクティブ操作の意義についての検討を行なう。

3.2 画像の構造化と高能率符号化

知的符号化の研究を開始するに当たり、画像の構造化と符号化の関係について検討を行った。いろいろな考え方があるが、仮に図 3.1 のように整理した。画像は理

第3章 インタラクティブ知的符号化の基本構成

解・記述するための「構造化」と通信・蓄積のための「圧縮」との2つの観点から捉えることができる。画像を理解することにより、構造化を行なう事が出来、それにより構造的な記述が可能となる。また逆に、構造法則が与えられれば、それから画像信号波形を合成することができる。このような構造化を行なうと、一般には、記述情報は増加することになり、情報量の削減が課題となって来る。構造化を圧縮に組み込むためには、記述の情報量が減少する条件を追加する必要がある。顔動画像の知的符号化においては、送受に共通のモデルを持つことにより情報量の削減を図ってきた。構造化を自動化することは例えば Computer Vision (CV) 等においても問題点が多く、難解な事項の一つであった。これに情報量削減の条件が加わると一段と難解となる。そこで、これを特定の切り口から人間による解析を行い、法則を導くための基盤データを積み上げていく必要がある。これをサポートする一種の符号化用編集機を構築し、画像とその表現で意図したものとの関連づけを行い、手がかりを得て行くことがまず必要である。

符号化用構造化は処理に時間がかかり、編集機により人間の操作が加わると、実時間での符号化は難しい。この関係を図3.2に示す。そこでは、非実時間処理による超低ビットレート画像符号化技術を用いる応用が、重要となって来る。例として、マルチメディアの電子メール等がある。Internetは広帯域化がすすみ、Mosaicなど動画伝送も活発化しているが、サーバー間や片方向が主で、1対1通信はコストが膨大と

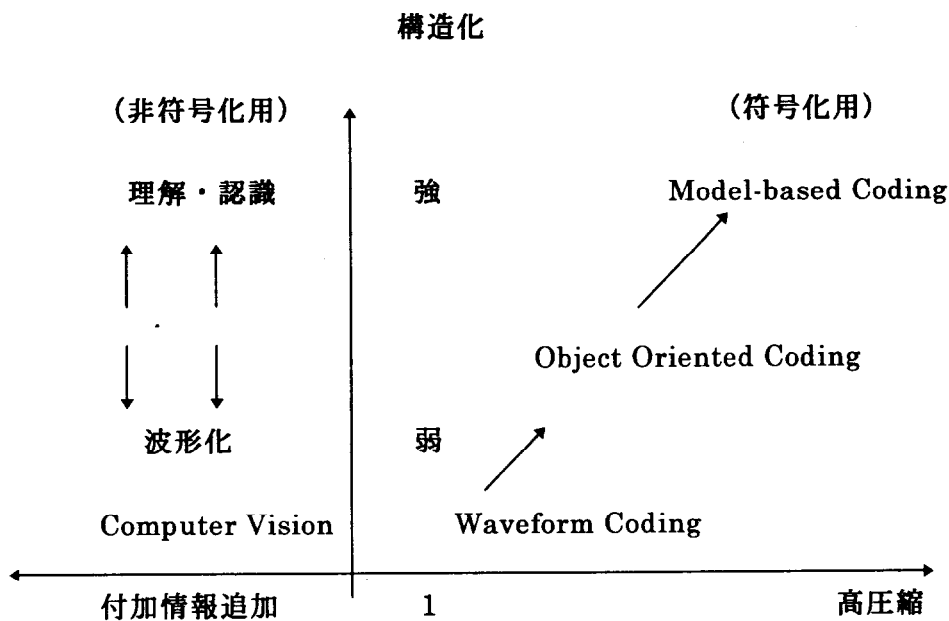


図3.1 構造化符号化

第3章 インタラクティブ知的符号化の基本構成

なることは避けられない。これら経費の増加で、現在のテキストメールの 50KB/mail のガイドラインは、今後も必要な制約事項と考えられる。電話並の料金で、マルチメディアメール通信を行うには超低ビットレートの符号化が必要となってくる。

応用として、テキストに加え、顔画像の符号化を行い、送信側で、編集機により、対話的に修正操作し、自分の意図する形状、拡張された表情の作成と低レートへの圧縮をともに実現する、記述システムの構築の実現が重要なものであると考えられる。

現在、電話、手紙などの古典的通信手段にかわり、数多くの新しい通信手段が使用され始めているが、その利用形態はそのメディア独特の非常に特化したものになっている。電子メール、ポケベル、似顔絵電子手帳などの使用形態は、その分野独特の進展がみられ、これを

Mediated Communication[30]

という特化した小世界での通信の慣習として分類する専門家もいる。電子メールに顔画像を加工添付するというシステムは単に技術のマルチメディア化だけでなく、新しい通信文化の形式を左右する重大な意味を持っているものでもある。

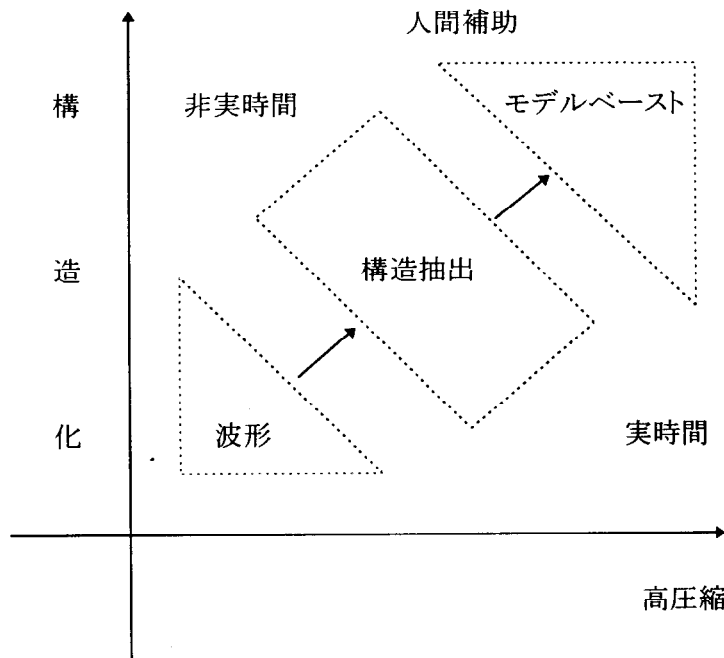


図 3.2 符号化における構造化と実時間処理性

Structuring and realtime processing in picture coding

第3章 インタラクティブ知的符号化の基本構成

以上の枠組みで、送信者の意図を表現することを、記述情報の効率化という条件を加えて実現していく。このための画像の「意図符号化編集機」を構築していくには、以下のような技術事項に関する研究を行う必要がある。

- ①入力の分析：顔形状検出、特徴点検出、追従
- ②モデル化：顔のワイヤーフレームモデル、多点化
- ③符号化：知的符号化
- ④合成・表示：復号画像表示
- ⑤編集：人間指示コマンドによる画像編集、拡張表現の対話的形成
- ⑥その他：対話編集機のソフトパッケージ化、Decoder ソフトの軽量化、P C 実験機上でのデモ

以上の対話型意図符号化編集機のブロック構成を図 3.3 に示す。

図 3.4 に画像符号化シンポジウム（P C S J）に発表された画像符号化・構造化関係の研究を技術の流れのマップにしたものを示す。86年に始まった分析合成符号化は特徴点の実時間追従、動き解析などに進み、構造化の取り扱い問題に到達した。また、顔表情の数値化、混合などの表情解析や、波形符号化的なアフィン変換動き補償（MC）による超低ビットレートの高効率符号化へも発展した。92年からは、3次元

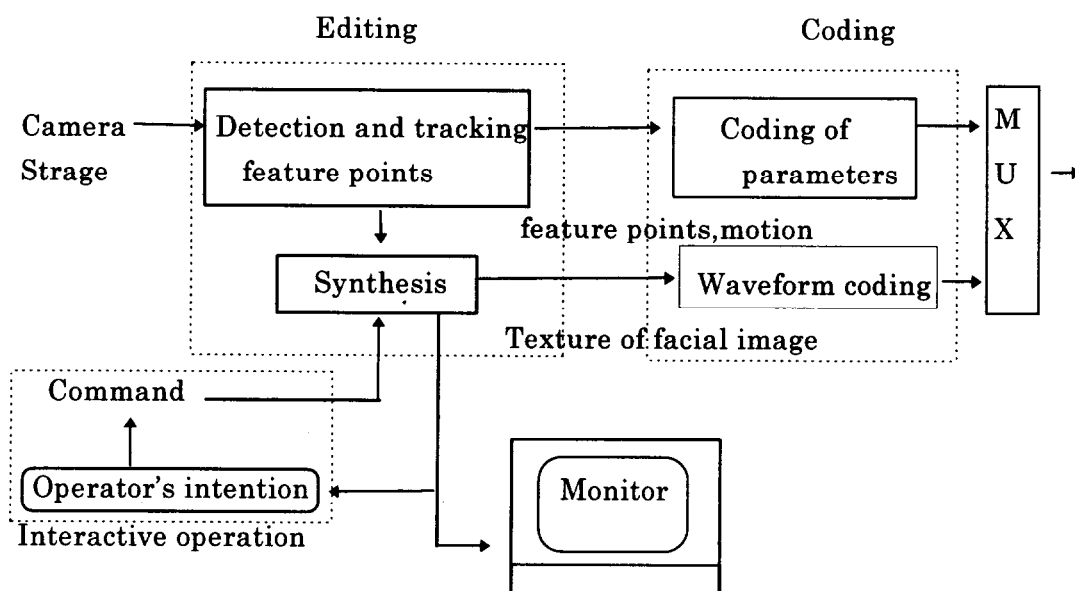


図 3.3 意図符号化編集機

Intention coding editor

第3章 インタラクティブ知的符号化の基本構成

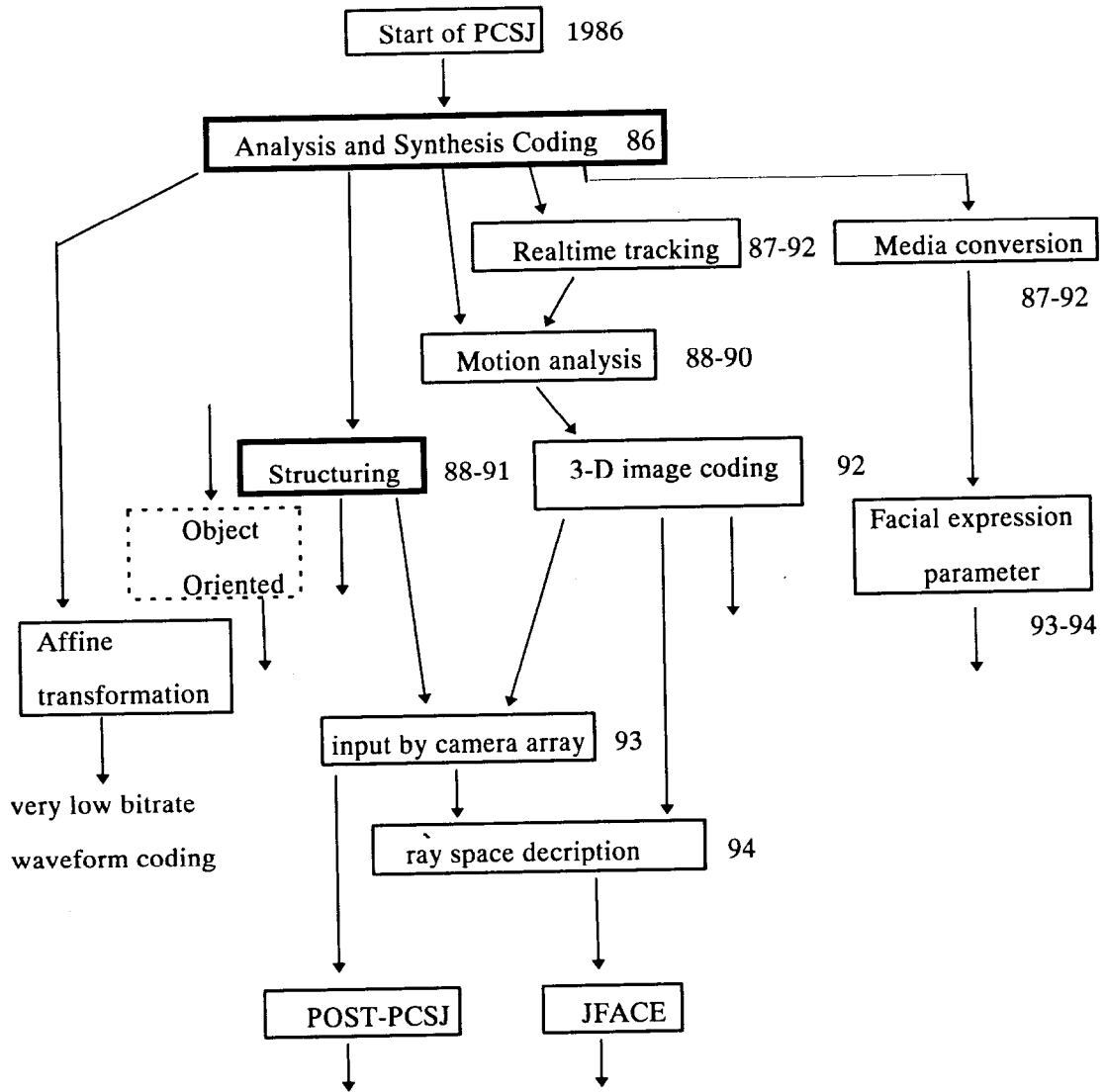


図 3.4 画像符号化シンポジウム(PCSJ)における画像符号化の研究経緯技術マップ
 Historical research diagram on picture coding from the papers
 in Picture Coding Symposium of Japan(PCSJ)

画像符号化が始まり、多眼カメラ入力、光線群空間記述、などに展開中である。一方海外では中間的な Object Oriented 方式の研究が進行している。

3.3 次世代符号化方式の動向

前節で述べた画像の構造化を動画像の高能率符号化に導入した次世代符号化方式の研究がなされている。本節ではその中から、分析・合成符号化（モデルベースト符号化）ないしは知的符号化についての概念をまとめ、次節で提案するインタラクティブ知的符号化の位置づけを明確化するための準備をしておく。

画像の超低ビットレート高能率符号化の研究動向については、文献[31]に述べられている。形状モデルを利用し、動き・形状変化情報を伝送するというモデルベースト符号化の原形については、1980年代初頭に先駆的研究が始まっている。顔動画像を対象としたモデルベース符号化の構想は、80年代半ば以降にテクスチャマッピング手法が導入され具体的な符号化方式として、研究が活発化した[32]。顔動画像を対象としたモデルベース符号化では、送・受信側双方で人物頭部の3次元構造モデルを共有し、この構造モデルに入力顔動画像中の顔部分の表情・動作と同一の表情・動作を与えるために必要な動き・形状変化に関わる情報のみを伝送することによって、顔動画像の伝送を実現しようとする点にある。テクスチャマッピング手法を用いて、原画像のテクスチャをワイヤフレームモデル上へ写像することにより、入力された顔画像のテクスチャが再生可能となる。送信側では入力画像を、知識に基づいて分析ないしは認識し、分析・認識結果をパラメータとして受信側に送信する。受信側では受信パラメータを基に、知識を用いて画像を合成して行く。

従来の波形符号化と分析・合成符号化等の違いを第0世代から第5世代に区分する考えがある[33][34][35][36]。これによると波形符号化を第0から第2世代に、第3世代に分析・合成符号化（モデルベースト符号化）、第4世代に認識・合成符号化（コマンド符号化）、第5世代に知的符号化を区分している。（表3.1参照）知的符号化は、画像の概念、意味あるいは画像が生成された意図や動機などを解釈・理解するものである。顔部品の動きの検出、例えば「まばたき」のようなものは、波形情報として送る方式、目のモデルパラメータとして送る方式（第3世代）、まばたきから、表情の理解や意図を解釈した場合は知的符号化（第5世代）として送る方式が可能で、知的符号化は概念の解釈や理解に人間が介在することができる場合は、符号化

第3章 インタラクティブ知的符号化の基本構成

表3. 1 画像符号化の世代区分 (文献[34]より引用)

世代区分	符号化方式	符号化に用いる知識	符号化レート (bps)	符号化方式の例
第0世代	直接符号化	振幅分布	$10^7 \sim 10^8$	標本値ごとのハフマン符号化など
第1世代	統計的冗長圧縮符号化	画素間相関などの統計的性質	$10^5 \sim 10^7$	予測符号化 変換符号化 ベクトル量子化など
第2世代	構造抽出符号化 特徴抽出符号化	動き情報・輪郭情報 などの特徴情報	$10^4 \sim 10^5$	輪郭符号化 構造・特徴情報に基づく適応符号化
第3世代	分析合成符号化 (モデルによる符号化) (パラメータ符号化)	画像のモデル 図形分離 特徴抽出	$10^3 \sim 10^4$	構造モデルに基づく 顔画像の分析合成符号化
第4世代	認識合成符号化 (コマンド符号化)	画像の認識と生成に 必要な知識と規則	$10^2 \sim 10^3$	
第5世代	知的符号化	画像の概念モデル 推論・試行 異種情報の統合	$10^1 \sim 10^2$	

方式の一部として実現可能である。

3.4 インタラクティブ知的符号化の提案

3.4.1 インタラクティブ知的符号化の基本構成

顔動画像の知的符号化は、送受信側双方において顔画像のモデルに関する共通の知識を備えることにより高能率な通信を可能とする方式として研究されて来た[29][37][38][39][41]。これまで初期画像が主に正面顔画像である場合を対象とした実験がなされているが、一般には送信側において画像の自動認識の問題を解決する点が課題となっている[39][40]。

本研究ではインタラクティブ知的符号化の概念を提案し[29]、従来知的符号化で問題となっていた画像認識処理を必須としない方式の有効性を検討して来た。グローバル動きに関し、2次元画面上のみでしか得られていない特徴点から、反復法を使わない直接計算法[42]により奥行きの変化分の検出を行い、その奥行き情報を用いて透視投影変換による3次元動き検出を行う新しい方式を開発した。従来、多変数の最適化により局所解に留まる可能性があったのに対し、最適解へ到達できるという特徴がある。また、全自動処理では検出が不安定になりやすい大きい回転動作を含む画像に対しても、破綻のない表現を確保することができる。

本研究では、インターネットのホームページのようなハイパーメディアを対象としたデータベースに顔動画像を蓄積しアクセスするような応用を考え、インタラクティブ知的符号化における諸技術を統合し、システム化を図ることを目指す。インタラクティブな操作は人間が補助支援することになるが、人間の操作や入力できる情報には限度があり、任意のデータを符号化装置に与えることができるものではない。例えば、人間が符号化に必要な回転角や平行移動量を具体的な数値で与えることは通常不可能なことである。より安定でより正確な結果が得られるような動き検出アルゴリズムを開発するために、どのようなインタラクティブ操作が人間にとって可能で、また符号化装置にとってどのような情報データが有用であるかを検討することが課題となる。インタラクティブ操作により、確定的な付加情報を与えることができ、従来の汎用で条件の与えられていない画像認識とは異なった、中間的な画像認識又は検出に問

第3章 インタラクティブ知的符号化の基本構成

題をブレイクダウンすることが出来る。難易度の低減を図ることによって、新たな中間的な画像認識又は検出の段階を設定することが可能になる。インタラクティブ知的符号化の枠組みの中で、頭部全体を剛体とみなした回転、平行移動に関するグローバル動き検出と、目、口等のローカルな動き検出との関係は、従来あまり論じられていなかったが、合成手法との関連で検討する必要がある。目、口等のローカルな動きの合成は FACS(Facial Action Coding System)[43]によりワイヤフレームモデルや筋肉モデルを駆動するものがほとんどである[44][45]。合成時には FACS におけるアクションユニットの線形結合で表現しようとする、現状の検出技術では回転等が伴う場合は誤差が生じやすい[46]。またマニュアルで行うとしても画面上での動きからアクションユニットの組を判定することは位置の指示などに比べると不確実性が高く格段に難しい。本研究では、口などのローカルな動きは非剛体の変形歪みも含めて頂点の位置で指示できるようにすることにより、大きい回転などのもとでも正確な口形の記述ができるような工夫を行った。また目などのように位置に拘わらず開閉度を判定しやすい顔部品に対しては、アクションユニットによる記述を行う構成をとった。

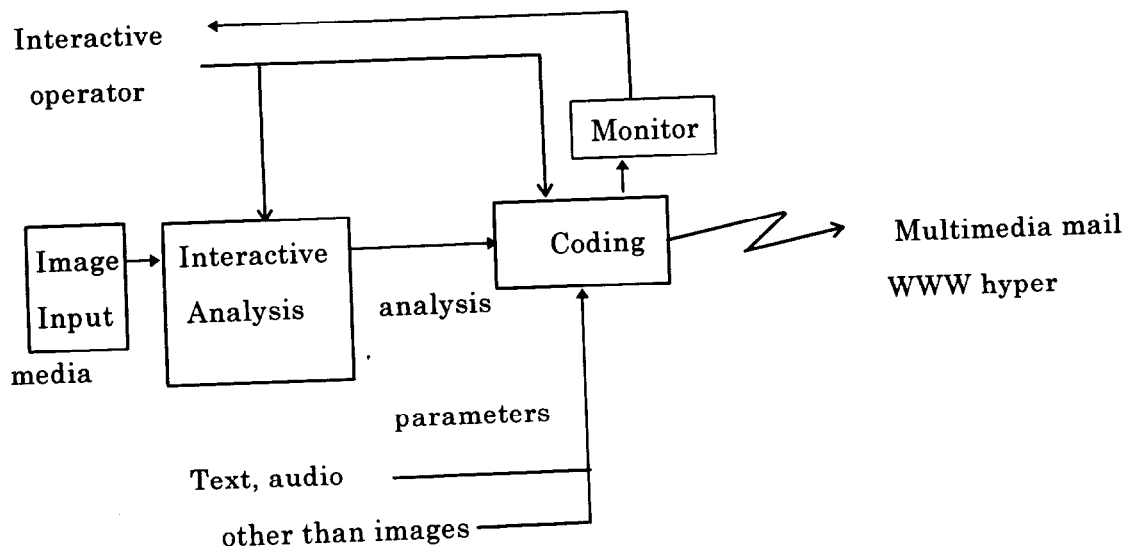


図 3.5 インタラクティブ符号化システム
Interactive Coding System

これを実現するためには、グローバル動き検出に基づいてワイヤフレームを変形した

後、移動後の状態で、特に口の動きを非剛体歪みと表情を含めたローカル動きとして検出している。また、合成においては、目のローカル動きを初期モデルで与え、口のローカル動きは移動後のモデル上で与えることにより非剛体の変形歪みや個人ごとに異なる動きも正確に記述可能となった。これにより、単眼カメラ入力で得られた実画像に追従したローカル動きを記述することが可能になった。以上のようなアルゴリズムをシステムとして統合し、符号化実験を行い、システムの検証を行った。

3.4.2 インタラクティブ性の意義

インタラクティブ知的符号化では、符号化パラメータの抽出や解釈・認識を人間が対話的に行なうが、このインタラクティブ知的符号化は単に人間が符号化の支援を行なうということだけではなく、以下のような積極的な意義を備えた方式と成りうるものである。

まず、入力画像に対してモデルのフィッティングや動きパラメータを検出するための特徴点の抽出、追従等をインタラクティブ操作により行い、自動検出では得られない高精度且つ自然感を再生し得る分析パラメータを獲得する。この分析パラメータを用いて画像の合成を行い（局所復号に相当する）、出力をモニタするという形で、結果をユーザ側に帰還する対話ループが形成されている。

インタラクティブ符号化の意義として

(a) 画像認識の汎用的な認識・動き検出のアルゴリズムに対する入力データの範囲を大幅に限定することにより、動きパラメータの検出を可能にする。

(b) 髭、眼鏡の処理など、従来自動化が難しいために、検討が行われていない方式に対し、アルゴリズムを開発し、知的符号化の性能向上を図る。

(c) 表情の加工・編集など送信者の意図を加える等の創造的処理を符号化と同時に行なうことができる。これにより、知的符号化の低ビットレート性を損なうことなく付加処理を行なうことができる。

などがあげられる。以下これら3点について詳説する。

(a) に関して：インタラクティブ操作により特徴点などの指示を行なうが、人間が符号化に必要な3次元動きパラメータ値を直接、瞬時に入力することは不可能であ

第3章 インタラクティブ知的符号化の基本構成

る。入力可能なものは、例えば2次元画面内の特徴点の位置などであるが、正確なパラメータを検出するための情報が最も多く、かつ最も指示が容易な方法を選ぶ必要がある。入力データは検出対象を大幅に限定することにより、動きパラメータの導出を確実に可能とするものを指示できる必要がある。あるいはまた検出結果を一定の誤差範囲内に限定することが可能な入力であることが必要である。従来ほぼ自動処理を前提になされてきた、動き解析の各種の方式には、未解決の問題も多い[45]。このような、効果的なインタラクティブ入力データにより、新たな検出方式を導くことも可能になってくる。

(b) に関して：従来の知的符号化では、画像自動認識に関する問題解決が前提条件であったため、これらは、それが解決したあとの問題として扱われ、十分に検討されていなかった。実際、現状では符号化の過程で自動検出処理を行なうのは困難であり、当面はインタラクティブ処理によってのみ可能と考えられる。この場合にも、最少の指示で最大の検出結果を得ることが課題となる。眼鏡のある顔画像から眼鏡のフレーム部を除去し、知的符号化による処理を施した後、復号結果に再び眼鏡のフレーム部の画像を付与することにより、眼鏡フレームに変形のない知的符号化が行なえることが報告されている。このような新しい処理を組込むことは、知的符号化の性能が向上したとみなすことができる。

(c) に関して：顔動画像の知的符号化において行われるワイヤフレームモデルのフィッティング結果を使用することにより、送信者の意図に応じて顔の表情の加工・編集を符号化の処理の中で行なえるという利点をもたらす。異なるワイヤフレームモデルを持つCGツールなどで表情加工の前処理を行なうと、ワイヤフレームモデルの不整合により、1つの動きが符号化段階で複数の表現になる可能性があり、符号化レートが増大する場合があります。知的符号化処理の中で行なうことにより、同一のワイヤフレームモデルの動きパラメータとなるため、レートの増加が最少になることが期待できる。なお、半自動の画像認識ツールとして、Cuginiらは“hint recognizer tool”の構想を述べ[47]、画像理解のシステムを構築しようとしたことがあった。本論文では、インタラクティブ操作を上のように検討し、具体的に動きパラメータの検出などのどのような情報を入力するのが良いかを調べ、符号化GUIツールとして開発して行く。

以上インタラクティブ符号化の意義について整理すると表3.2のようになる。まず、従来、顔動画像の諸認識や、特徴点検出・追従が完全には行なえない場合があり、汎用的な全自動化ができなかった。しかしある条件を加えると汎用とマニュアルの中間

第3章 インタラクティブ知的符号化の基本構成

的な検出の問題、つまり条件付きの可解な問題に転換することができるようになる。次に、インタラクティブ知的符号化が符号化を実行出来るようになると、前処理としての画像認識などの問題に拘束されることなく、知的符号化本来の性能向上や高機能化に専念できるようになり、知的符号化そのものの研究や実用化の推進が可能となる。更に、インタラクティブ知的符号化のインタラクティブ操作の段階で得られる構造モデルや動き等のパラメータは単に符号化に使用されるだけでなく、送受信において、画像の加工・編集などの創造的処理を可能とするため、符号化からその周辺へ拡大した信号処理への展開ができるようになって来る。

以下の章では、4章に於いて、上記(a)に述べたようにインタラクティブ操作を導入した新たな中間的動きパラメータ検出アルゴリズムについて述べる。4、5章で(b)に述べた知的符号化全体を構築し、眼鏡画像処理という高機能化を与える処理について述べる。更に、5、6章では、インタラクティブ操作の関連した加工・編集処理である2人の画像からテクスチャやワイヤフレームモデル、動きパラメータ等を入替え新たな画像を合成する実験を示して行く。

表 3.2 インタラクティブ操作の意義

(a)	画像認識、特徴検出の新たな 中間的な段階 を設定し、対象の限定を行なう。
(b)	知的符号化符号化の画像認識部以外の 高性能化、高機能化 を図る先行的アルゴリズムの開発が行なえる。 ——髭、眼鏡画像処理、背景処理など
(c)	表情の 加工・編集 など操作者の意図を反映した創造的処理が可能となる。 送受を問わず、画像テクスチャや動きを入替えたり、加工することが可能となる。

3.5 むすび

本章では、1980年代から研究が行なわれてきた、動画像のモデルベースト符号化、知的符号化の研究動向を整理し、新たに「インタラクティブ知的符号化」の基本構成を提案した。これは第3世代から第5世代にまで及ぶ符号化方式であり、インタラクティブ性を導入することにより、従来の研究の一部が加速されるという効果をもたらす。また、インタラクティブ性の意義について論じ、単に人間が符号化の支援を

第3章 インタラクティブ知的符号化の基本構成

行なうということだけでは無く、積極的な意義を備えた方式と成りうるものであることを述べた。

応用としては、MM-email 用、「意図・符号化・編集機」の研究を分析合成符号化に人間の補助を対話的に付加システムを想定している。具体的には、従来の分析合成符号化システムを統合し、これに、意図を画像に反映させる対話型編集システムを構築していく。

以下では、第4章では、動き検出の分野で、頭部全体の動きを検出するための新しい「グローバル動き検出方式」、「インタラクティブ操作に起因する新しいノイズ除去フィルタシステムとその検出精度」、第5章では、「インタラクティブ知的符号化に対応した新たなローカル動き記述方式とその検出精度」、「眼鏡のある顔画像に対する処理と上書き処理操作の精度評価」、第6章では「インタラクティブ操作の関連した加工・編集処理である2人の画像からテクスチャやワイヤフレームモデル、動きパラメータ等を入替え、新たな画像を合成する実験を示し、インターネットのホームページを用いた転送実験と再合成画像の主観評価実験」などについて述べる。

第4章 グローバル動き検出とノイズ除去フィルタ

4.1 まえがき

本章ではインタラクティブ知的符号化の基本構成に基づいて、具体的な符号化システムを構築して行く。顔動画像のインタラクティブ知的符号化の構成においてはまず、頭部全体の3次元の動きパラメータを正確に検出する必要がある。このため、動画像系列から3次元の動きパラメータを検出するアルゴリズムの調査を行なう。それによると一般的な解法は無いため、顔動画像における3次元動きパラメータの検出という検出対象の限定を図ることによって、どのような条件が得られ、それをどのように検出アルゴリズムに反映していくかの検討を行う。顔動画像に対する3次元動きパラメータの検出という検出対象の限定として考えられる事項には、頭部を剛体と仮定した時に、顔部品を動き検出の手がかりとして用いることや、顔動画像の撮影においては、カメラから撮影対象の人物までの距離はある範囲にあるなど、各種のものがあろう。本章では、顔部品の中から手がかりとして用いる特徴点の選定、それらを用いて最小2乗法により3次元動きパラメータを検出するアルゴリズムの開発、インタラクティブ操作により入力される特徴点に重畳するノイズの除去を行なうフィルタの構成、更にインタラクティブ操作により入力される特徴点の精度を検証し、インタラクティブ知的符号化システムとして妥当なシステムが構成されるものであることなどを検討して行く。

まず、前半において、3次元のグローバル動きパラメータ検出を行なう新たなアルゴリズムについて述べ、後半において、インタラクティブ操作により重畳するノイズの除去を行なう新たな構造的フィルタの提案と精度の検証を行なう[48][49]。

4.2 パラメータの検出のシステム構成

図 4.1 は、本章で提案するインタラクティブ知的符号化システムの構成例である[48]。この図はインターネットのホームページのようなハイパーメディアを対象としたデータベース（以下 HMDB と略す）に顔動画像を蓄積しアクセスするような応用

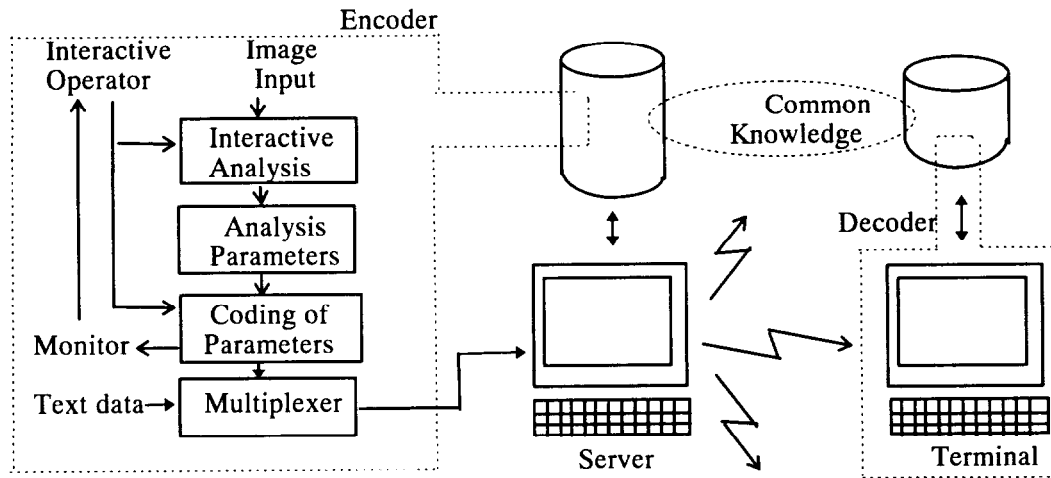


図 4.1 HMDB に対する顔画像のインタラクティブ知的符号化システム

Interactive model-based coding system for hypermedia database.

を想定した場合の構成図である。インタラクティブに符号化されたデータは HMDB のサーバに蓄積され、リモートの端末からのアクセスにより読み出される。送信側と共有する知識は復号プログラムの処理手順に内蔵されている。図 4.2(a),(b)はその符号化、復号部のブロック図である。入力画像から 3 階層の画面の作成、ワイヤフレームモデルのフィッティング、2 次元特徴点の検出、ローカル動きの検出が行われる。破線はインタラクティブ操作のある処理、実線は自動処理アルゴリズムを示す。まず、階層化部では、入力フレームから頭部、首から下の肩画像、及び背景の 3 つの画面を作成する。

まず、このインタラクティブ知的符号化システムにおける 3 次元動きパラメータ検出アルゴリズムについて検討を行なって行く。従来の 3 次元動きパラメータ検出においては、非線形な連立方程式を反復法により求めて行くことがなされていた。インタラクティブ知的符号化に対応した新しい動きパラメータ検出アルゴリズムを開発して行く。インタラクティブ操作により得られる有効な情報を生かし、またこのような応用において求めるパラメータの存在範囲が限定されてくるという条件を組込み、線形な検出アルゴリズムを導出することを行なって行く。3 次元空間内において線分の長さが回転と平行移動にたいして不変であることを用いて、3 次元の奥行き方向のパラメータを第一に求める。これには顔部品上の特徴点の距離が用いられる。これにより得られる条件式の数は未知数の数よりも多くなり、従って最小 2 乗法により安定な解

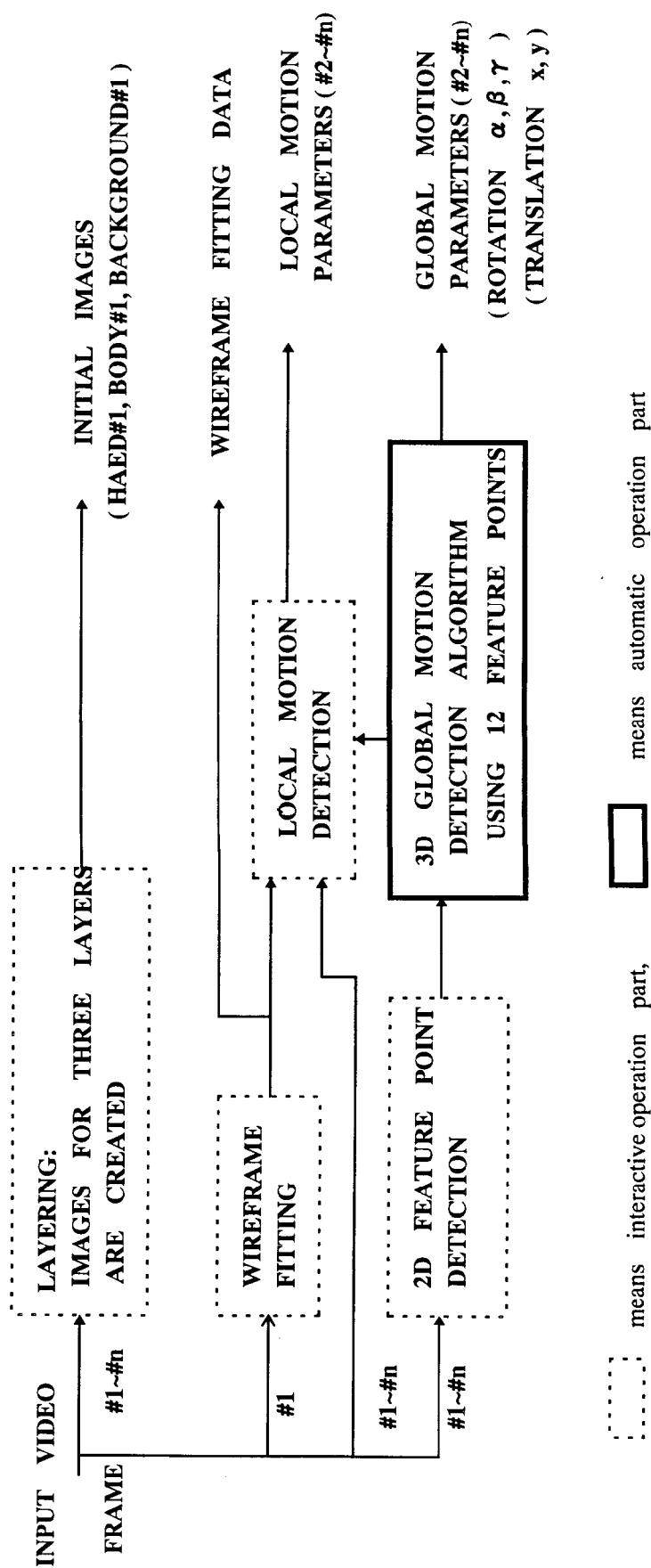


図4.2(a) インタラクティブ知的符号化の符号化部
Encoder of interactive model-based coding

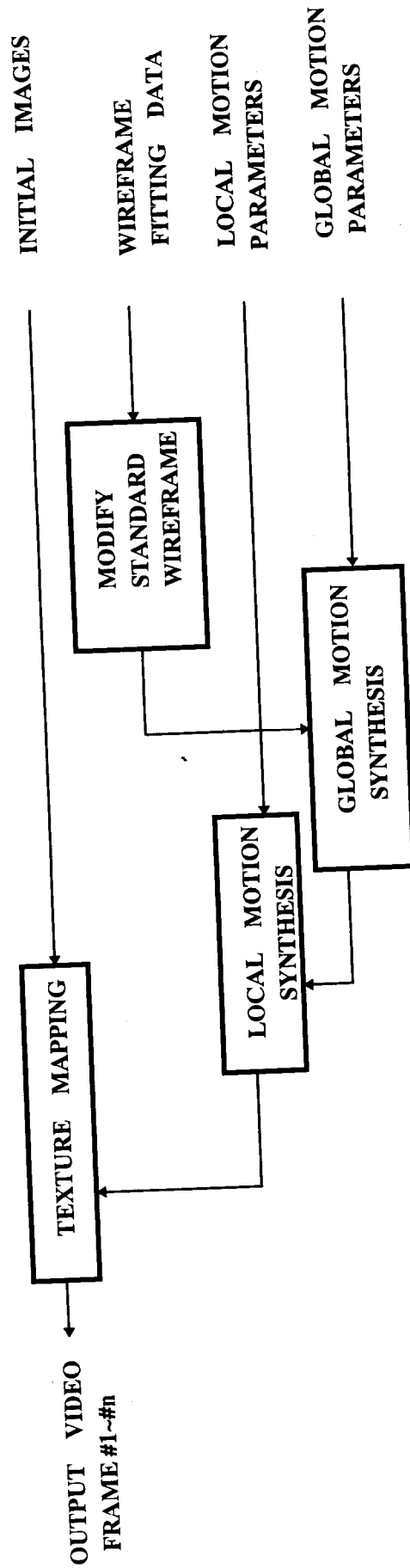


図 4. 2(b) インタラクティブ知的符号化の復号部
Decoder of interactive model-based coding

を求めることが出来る。以下、この検出アルゴリズムについて詳細に述べて行く。

顔動画像のモデルベースト符号化における動きパラメータ検出では、通常2次元の画像系列から処理がなされる。つまり、2次元情報から3次元の動きパラメータを検出する必要があるため、一般的で汎用な解放を求めることは難しい。インタラクティブ操作によって、より好都合な信頼性の高い特徴点の位置情報が供給される。又、対象としている応用の環境条件により、求める動きパラメータの存在範囲が限定でき、解法の範囲を限定して行くことが出来る。顔画像のもとになる頭部は一般には非剛体である。頭部の動きは剛体を仮定したグローバル動きと非剛体を想定したローカル動きに分解することが出来る[45][50]。ローカル動きは目、口などの顔部品の動きとして表現出来る。本論文ではグローバル動きは3次元の剛体の動きパラメータ検出問題として扱い、ローカル動き検出はグローバル動きパラメータが検出された後、それに加えられる動きとして検出することになっている。

特徴点の対応関係から動きを検出する手法に関して、文献[51]に数多くのアルゴリズムが整理・分類されている。表4.1に全体を6つの分野に分類して示す。すなわち、用いる入力として3-D, 2-D情報の違いがあり、検出する動きパラメータの次元として3-D, 2-Dがあり、組み合わせとして3通りが考えられている。更にそれらの各々について、入力の情報が座標値として与えられる点对応の場合と、ある直線が別の直線と対応している線対応の場合がある。顔の表面は曲面であり、線対応の対応づけは行いにくい。従って、本章で扱うグローバル動きパラメータ検出アルゴリズム

表 4.1 動き検出アルゴリズムの分類

([51]T.S.Huang,1994)

入力条件	検出パラメータ
3-D 線対応	3-D
3-D 点对応	3-D
2-D 線対応 3-D	
2-D 点对応 3-D	
2-D 線対応 2-D	
2-D 点对応 2-D	

ムは、このうち2-Dから3-Dの点对応に分類されることになる。

線形な解法として Tsai らは、法線の関係を用いて奥行き方向の項を消去することにより、対応する問題を解いている[52]。しかしながら、この手法はノイズに敏感であり、非線形の反復手法の方が J.Weng らにより推奨されている[53]。又、上記文献[51]の IV の B.の -1) によると、連立方程式系が正則という条件があれば、単一の解が得られることが示されている。しかし正則となる条件が求められていない。文献[51]にあげられている2-Dから3-Dの点对応の

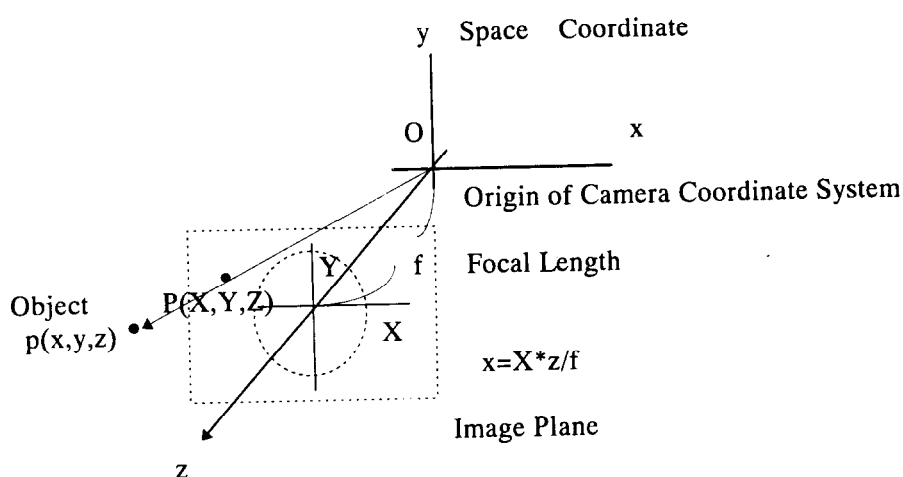


図 4.3 カメラ座標(x,y,z)と対応する画像の撮像面における座標(X,Y,Z)の関係

Coordinate system(x,y,z) attached to the camera,
and the corresponding image coordinates(X,Y).

The center of rotation is on image plane

アルゴリズムは動きパラメータ検出のために、反復処理を行なっている。

本章では、これらの従来研究技術を踏まえ、新たに反復手法を必要としない線形な解法を開発して行く。

4.3 3次元パラメータの検出

図 4.3 は文献[54]に基づき、カメラ座標(x,y,z)と対応する画像の撮像面における座標(X,Y,Z)の関係を示している。画像面の点Pは3次元空間内の物体の点pの透視投影である。インタラクティブ知的符号化の初期画像（第1フレーム）に対してワイヤ

フレームのフィッティングがなされているものとする。3次元空間内での第1フレームにおける特徴点の座標を (x,y,z) とする。

この時、

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} R \\ t \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (4.1)$$

但し、 (x',y',z') は第1フレームを基準にして移動した後の座標値で、第2フレーム以降の特徴点の座標値を表す。

$R=(r_{ij})$, $1 \leq i, j \leq 3$ は回転の変換を表す行列、 $t=(t_1,t_2,t_3)^t$ は平行移動の変換項を表す。画像の撮像面はカメラの原点 O から焦点距離 f の位置にある。空間座標 (x,y,z) とカメラ座標 (X,Y,Z) には、透視投影の関係から、

$$x = \frac{zX}{f}, \quad y = \frac{zY}{f} \quad (4.2)$$

の関係がある。回転の中心は画像の撮像面にあると仮定されている。第1フレーム上の X, Y と第2以降のフレーム上にある X', Y' の4個のデータがインタラクティブ操作で与えられる。又、第1フレームの奥行き Z は3次元ワイヤフレームモデル上の標準データとして与えられるものとする。第2フレームの z' は未知数である。(4.2)式の関係を用いて(4.1)式を観測データに置き換えると超越方程式となる。これはたとえ反復法で解を探索しても、ユニーク性が得られないことから、実際に最適解を求めることは困難な場合が多い[51]。そこでまず未知数 z' を求めるため、3次元空間内で距離が不変であること

$$\|p_i - p_j\| = \|p'_i - p'_j\| \quad 1 \leq i, j \leq N \quad (4.3)$$

を使用することにする[50]。しかし一般には2次の

9個以上の連立方程式となり、複数次不定解の数(例えば512個など)や、全ての実数の解は求めることができない[50]。ここでは N 個の特徴点に関する第1、第2フレームの座標のうち既知の X, Y, Z, X', Y' を用いて z' を定める式を求める。

4.3.1 奥行き z の検出[42]

グローバル動きパラメータを求める第一段階として、奥行き z' を最小2乗法により求める。

複数の特徴点にインデックス $i(i=1, \dots, N)$ を付して表すと、(4.3)より

$$\begin{aligned} & (x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2 \\ & = (x_i'-x_j')^2+(y_i'-y_j')^2+(z_i'-z_j')^2 \end{aligned} \quad (4.4)$$

(4.2)式の関係から

$$\begin{aligned} & (z_iX_i-z_jX_j)^2+(z_iY_i-z_jY_j)^2+F^2(z_i-z_j)^2 \\ & = (z_i'X_i'-z_j'X_j')^2+(z_i'Y_i'-z_j'Y_j')^2+F^2(z_i'-z_j')^2 \end{aligned}$$

これは、(4.2)式の関係を用いて、

$$\begin{aligned} & (z_iX_i-z_jX_j)^2+(z_iY_i-z_jY_j)^2+f^2(z_i-z_j)^2 \\ & = (z_i'X_i'-z_j'X_j')^2+(z_i'Y_i'-z_j'Y_j')^2+f^2(z_i'-z_j')^2 \end{aligned} \quad (4.5)$$

これは、

$$A_{ij} z_i'^2+B_{ij} z_i'z_j'+C_{ij} z_j'^2=D_{ij} \quad (4.6)$$

の形になっている。但し、

$i=1,\dots,N$

$$A_{ij} = X_i'^2+Y_i'^2+f^2, \quad C_{ij} = X_j'^2+Y_j'^2+f^2 \quad (4.7)$$

$$B_{ij} = -2(X_i'X_j'+Y_i'Y_j'+f^2) \quad (4.8)$$

$$D_{ij} = (z_iX_i-z_jX_j)^2+(z_iY_i-z_jY_j)^2+f^2(z_i-z_j)^2 \quad (4.9)$$

と置き換えるものとする。

(4.6)は、 $N \geq 3$ で方程式の数は未知数以上になる。ここで、微少な差分 dz' を導入し、

$$z'=z+dz'$$

と置くと、以下のようにして未知数 z_i', z_j' を求めることが出来る。

まず、 $A_{ij}=A, B_{ij}=B, C_{ij}=C, D_{ij}=D$ と置き直し、

$$z_1'=z_1+dz_1'$$

$$z_2'=z_2+dz_2'$$

とすると、(4.6)は

$$A(z_1+dz_1')^2+B(z_1+dz_1')(z_2+dz_2')+C(z_2+dz_2')^2=D \quad (4.10)$$

となる。両辺を Z_a^2 で割り、 $dz_1'/Z_a=w_1, dz_2'/Z_a=w_2$ と置くと

$$\begin{aligned} & A\{(z_1/Z_a)^2+2(z_1/Z_a)w_1+w_1^2\} \\ & \quad +B\{z_1z_2/Z_a^2+z_1/Z_a w_2+z_2/Z_a w_1+w_1w_2\} \\ & \quad +C\{(z_2/Z_a)^2+z_2/Z_a w_2+w_2^2\}=D/Z_a^2 \end{aligned} \quad (4.11)$$

となる。ここで Z_a としては z からの距離の平均を用いることにする。又、現在考えている撮影条件においては、 z_i や dz_i として、例えば

$$z_i \doteq 200\text{cm}$$

$$dz_i \doteq 5\text{cm}$$

と仮定することが可能であり、従って、

$$w_i = dz_i / Z_a \leq 0.025$$

$$w_i^2 = (dz_i / Z_a)^2 \leq 0.000625$$

と見積もることが可能である。

これを、用いて、2乗項を無視すると、

$$\begin{aligned} & \{2A(z_1/Z_a) + b(z_2/Z_a)\}w_1 + \{2C(z_2/Z_a) + B(z_1/Z_a)\}w_2 \\ & = D/Z_a^2 - A(z_1/Z_a)^2 - Bz_1z_2/Z_a^2 - C(z_2/Z_a)^2 \end{aligned} \quad (4.12)$$

なる線形連立方程式が得られる。

2個の特徴点の組に対し、1個の線形方程式を得ることが出来る。求めるべき未知数は奥行き z_i であるから、その数は特徴点の個数に等しい。又、3点 p_1, p_2, p_3 に対し、(4.12)から3個の独立な方程式を得ることが出来る。N個の特徴点に対しては、 $N C_2$ 個の独立な方程式を得ることが出来る。特徴点が12個の場合、 $12 C_2 = 66$ 個の独立な方程式を得ることが出来、最小2乗法によって安定な解を得ることが出来るようになる。(4.12)から求まる、 w_i に対し、 Z_a を乗じれば z_i の差分が dz_i えられる。以上がグローバル動きパラメータを求めるための奥行きの検出手法である。

4.3.2 3次元グローバル動きパラメータの検出

次に、4.3.1 で求めた情報を用いて、実際にグローバル動きパラメータを求める。

z_i, z_j が求まると、2次元座標値 X', Y' から3次元座標値 (x', y', z') を求めることが出来る。既知となったデータ $x_i', y_i', x_i, y_i, z_i$ を(4.1)式に代入する。この時、8個の未知数

$$r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, t_1, t_2$$

に関する、N個の線形な方程式が得られる。最小2乗法[55][56]により、人物の頭部を剛体仮定することにより、上記8個のグローバル動きパラメータを求めることが出来る。最小2乗法の適用における未知数の関係は、

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 & 0 \\ r_{21} & r_{22} & r_{23} & 0 & t_2 \end{bmatrix} \begin{bmatrix} x & y & z & 1 & 1 \end{bmatrix} t \quad (4.13)$$

1個の特徴点の対応 $(x_i, y_i, z_i) \Leftrightarrow (x_i', y_i', z_i')$ に対して、(4.13)から2個の条件式を得ることが出来る。従って、N個の特徴点に対しては2N個の条件式が下記のような

式(4.14)で得られることになる。

$$\begin{bmatrix} x_1' & x_2' & \cdots & x_N' \\ y_1' & y_2' & \cdots & y_N' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 & 0 \\ r_{21} & r_{22} & r_{23} & 0 & t_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ y_1 & y_2 & \cdots & y_N \\ z_1 & z_2 & \cdots & z_N \\ 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad (4.14)$$

4.3.3 近似を行なう場合

前節では透視投影法に基づいた3次元のグローバル動きパラメータを求めたが、ここで、比較のため平行投影に似た近似を行なう場合についてもグローバル動きパラメータを検出するアルゴリズムを求めておく。これは、透視投影の場合に比べ、近似されたグローバル動きパラメータが求められることになる。

(4.2)を(4.1)に代入することにより、

$$\begin{pmatrix} \frac{z_i' X_i'}{f} \\ \frac{z_i' Y_i'}{f} \\ z_i' \end{pmatrix} = \begin{pmatrix} & & \\ & R & \\ & & \end{pmatrix} \begin{pmatrix} \frac{z_i X_i}{f} \\ \frac{z_i Y_i}{f} \\ z_i \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (4.15)$$

すなわち、

$$\begin{pmatrix} X_i' \\ Y_i' \\ f \end{pmatrix} = \frac{z_i'}{z_i} \begin{pmatrix} & & \\ & R & \\ & & \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ f \end{pmatrix} + \frac{f}{z_i} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (4.16)$$

を得る。(4.16)式の未知数は Z_i' と r_{ij} と t_1, t_2, t_3 であるが、最下行は合成時の座標の奥行きであるため、画像合成には必要ない。従って、求めるべき未知数は9個である。ここで、回転変換を表す行列 R は Euler の3つの角度である3個のパラメータで表現出来るため、本質的に必要な未知数の個数は6個となるが、ここでは形式上

8個にして、超越関数方程式になるのを防ぐ近似を行なう。

通常のカメラ撮影の状況で、 Z_i に対する、 Z_i' の変動は比較的小さいと考えられ、 $z_i/z_i'=1$ と見做しても誤差は小さいと考えられる。そこで、 $z_i/z_i'=1$ とおくと

$$\begin{pmatrix} X_i' \\ Y_i' \\ f \end{pmatrix} = \begin{pmatrix} R' \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ f \end{pmatrix} + \frac{f}{z_i} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (4.17)$$

となり、8個の未知数に対する連立一次方程式系となる。顔画面上の適当な個数の特徴点を第一フレームと符号化フレームで指定すれば、最小2乗法等で回転のパラメータを求めることが出来る。このパラメータは直接軸回りの回転量等に対応するとは限らないが、顔全体を剛体とみなして回転させる目的にはこのパラメータ状態でのデータで十分である。この式は回転変換の行列 R の中に近似で消えた z_i / z_i' の平均値が残るため、純粹の回転の他に Z 方向の平均的動きまで表現できると考えられる。更に回転変換の行列 R の未知数を3個に削減すると z 方向のシフトを表現できないことによる誤差が発生すると考えられる。

ここで従来の平行投影の場合をこの近似方式との比較をするために、述べておく。

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} 1 & \phi & -\psi \\ -\phi & 1 & \theta \\ \psi & -\theta & 1 \end{bmatrix} \begin{bmatrix} X_i' \\ Y_i' \\ Z_i' \end{bmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (4.18)$$

上記の近似に比べ、スケールファクタが無い場合、奥行きに大きい変化のある動きがあった場合には、グローバル動きパラメータ検出はエラーが発生する。この現象は次節の実験により示される。

以上、グローバル動きパラメータ検出のアルゴリズムについて述べて来た。単眼カメラからの動画像入力を想定すると、顔動画像のグローバル動き検出を行うためのインタラクティブ操作は、表 4.2 のように分類できる。2次元画面上での参照点の座標位置を指示入力する場合について、その点の数により分類している。グローバル動きパラメータは、回転3個 (α, β, γ) と平行移動2個 ($\Delta x, \Delta y$) である。零個はインタラクティブ操作がない場合で、全自動検出に対応する。2次元座標値として1点につき X, Y 座標の2つの条件が得られる。3点で条件数は未知数の個数を上回るが、指示点座標値には雑音が加わるため、連立方程式による解法では安定な解は得られな

い。一般に最小2乗法を用いる際には、未知数の個数の2乗分の個数の条件が適当とされており[56]、この問題に対しては12点の特徴点を用いることが妥当であることが分かる。顔部品の頂点のうち、目の端点は横一直線上にあり、これだけでは動き検出のための参照情報として不十分である。本論文では図4.4に示すように、頭部全体の広い範囲から12個の参照用の特徴点を用いた。

なお、本方式では、2次元画面上の点から3次元動きパラメータを検出するための工夫を行っていることを特徴としており、従来のモーフィング[57]などのように、単に前後のフレームから中間フレームを内挿するだけの方式とは、本質的に異なる。

表4.2. 顔画像のインタラクティブ グローバル動きパラメータ

($\alpha, \beta, \gamma, \Delta x, \Delta y$) の検出のために使用する情報と対応アルゴリズムの関係

Table. Relation between the number of points for global motion parameters detection and corresponding algorithm

インタラクティブ操作 による入力点の数	アルゴリズム
0 (個)	全自動未知アルゴリズム
1 (2条件)	半自動未知アルゴリズム
2 (4条件)	"
3 (6条件)	限界アルゴリズム(連立方程式) (入力点は誤差なしであること)
4 (8条件)	最小2乗アルゴリズム
•	• • • • •
12 (24条件)	安定最小2乗アルゴリズム (入力点の誤差を許容)
•	• • • • •
$\alpha, \beta, \gamma, \Delta x, \Delta y$ (解の入力)	アルゴリズムなし

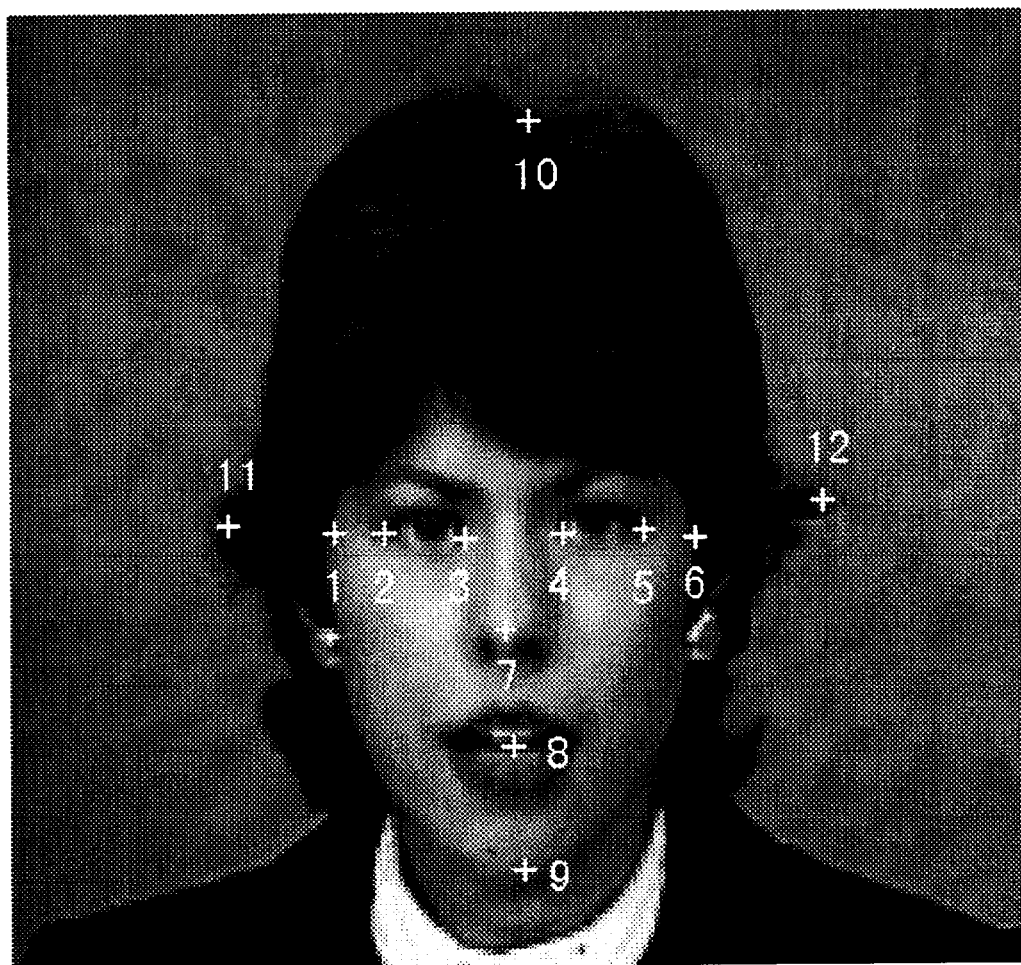


図 4.4 特徴点の配置図

Arrangement of feature points

4.4 グローバル動きパラメータ検出の基本実験

これまでの節で述べたグローバル動きパラメータ検出アルゴリズムの基本性能を評価するため、CCITT（現在の ITU-T）の標準画像を用いてグローバル動きパラメータの検出を行った。従来、上記近似方式のような平行投影に基づいたグローバル動きパラメータ検出がなされていたが、比較的大きい動きに対しては、検出エラーが大きくなる問題があった。平行投影のまま、大きい動きに対処するため、第一段階で平

行移動成分を検出し、検出された平行移動成分を差し引き方式が検討されていた[58]。又、反復法により、奥行きを検出を行い、近似誤差を低減するほうしきなども検討されている[53]。しかしながら、いずれにしても奥行きを検出を行わないことにより、動きによっては動きパラメータに誤差成分が残るといった問題があった。ここでは、4.3.1 で述べたアルゴリズムにおいて、特徴点の数を $N=12$ としてグローバル動きパラメータの検出をおこなう。第一段階で奥行き差分 dz' を求める。第二段階で、 $2N=24$ 個の条件式を用い、最小2乗法によって回転と平行移動のグローバル動きパラメータを求める。検出アルゴリズムの動作を調べるため、いくつかのテストデータを入力した。このテストデータは第一フレームと符号化フレームの奥行き差分が予め分かっている。検出アルゴリズムは第一フレームの3次元の情報と、符号化フレームの2次元情報をもとに3次元のグローバル動きパラメータを検出する。図4.5(a) は検出された奥行き z の差分 dz_det が正しい値 dz_ref と良く一致している状態を示している。図4.5(b) は第一フレームと符号化フレームとの奥行き z の差分 dz を示す。入力はCCITTのClaireである。この差分 dz はClaireの動きに対応している。このようにして求めた奥行き情報をもとに2次元対応から得られるものよりも正確な3次元の透視投影情報が得られる。

4.3.1 のグローバル動きパラメータ検出と4.3.2 の近似方式により動きパラメータ検出を行った。検出されたグローバル動きパラメータを用いて第一フレームのワイヤフレームモデルを変形合成したものの動きは全体として滑らかであった。図4.6(a) は4.3.2 の近似方式により検出されたグローバル動きパラメータにより合成された第10フレームのワイヤフレームである。図4.6(b) は4.3.1 の近似方式により検出されたグローバル動きパラメータにより合成された第10フレームのワイヤフレームである。どちらも画像は原画像をそのまま用いている。図4.6(a) では、顔のワイヤフレームモデルが実際より縦長になっている。一方、図4.6(b) では顔のワイヤフレームモデルは実画像によく一致している。

図4.7(a), (b) は4.3.1 のグローバル動きパラメータ検出アルゴリズムにより検出されたグローバル動きパラメータを用い、目と口の動きを最大数個程度の座標値で表すというローカル動きパラメータで記述した場合の合成画像の例である。図4.7(a)はCCITTのClaireの第94フレーム、図4.7(b)は第112フレームである。

表4.3にここで行った符号化における所要ビット数を示す。主にグローバル動きパラメータ検出の動作を総合的に確認することがもくてきであるため、ローカル動きパラメータは簡易な方式によった。又、初期画像はQCIFサイズのもの1枚を用いて

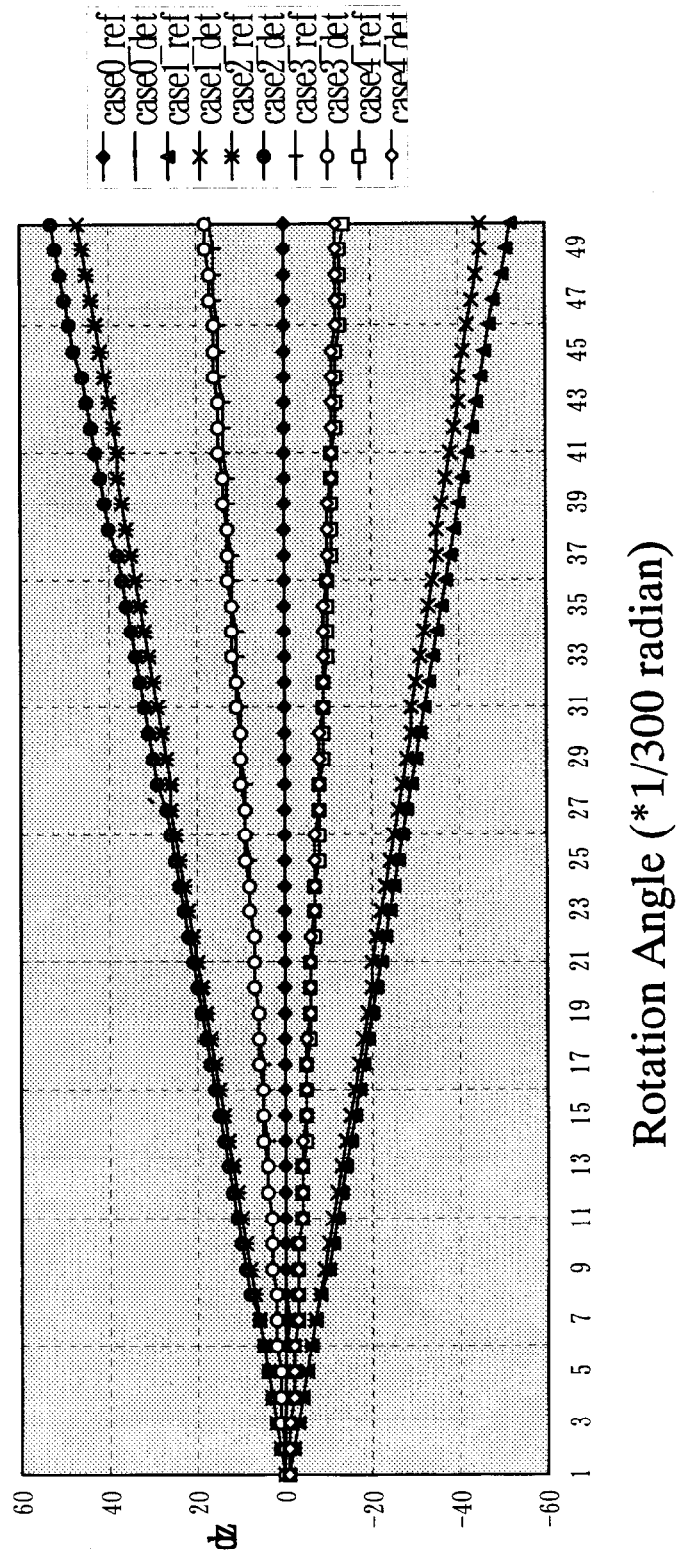
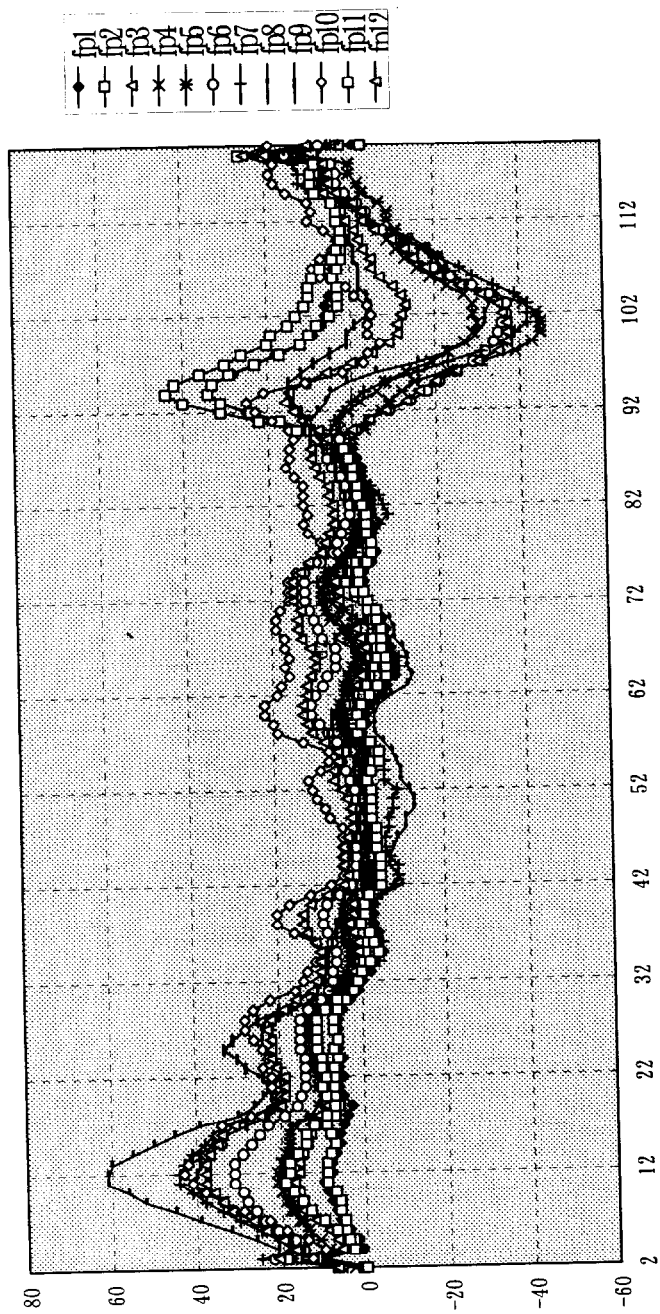


Fig.4.5(a) Comparison Test of Depth Values for Artificial Motion Data



Frame Number (Claire Sequence)
Fig.4.5(b) Detected Depth Parameters for Claire Sequence(FP_Z.XLS)



図 4.6(a) 4.3.2 の近似方式によるグローバル動きパラメータで合成したワイヤフレームモデル(CCITT Claire, 第 10 フレーム)

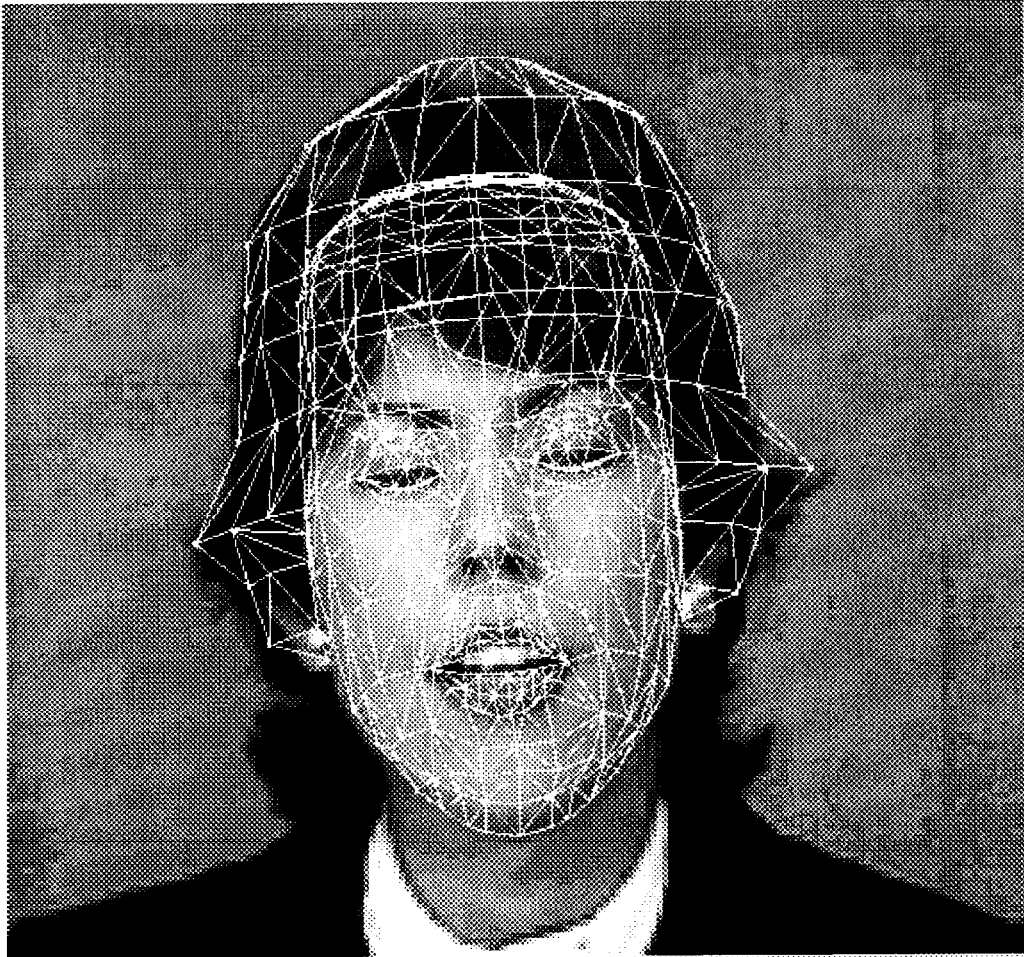


図 4.6(b) 4.3.1 の透視投影の検出方式によるグローバル動きパラメータで合成したワイヤフレームモデル(CCITT Claire, 第10フレーム)

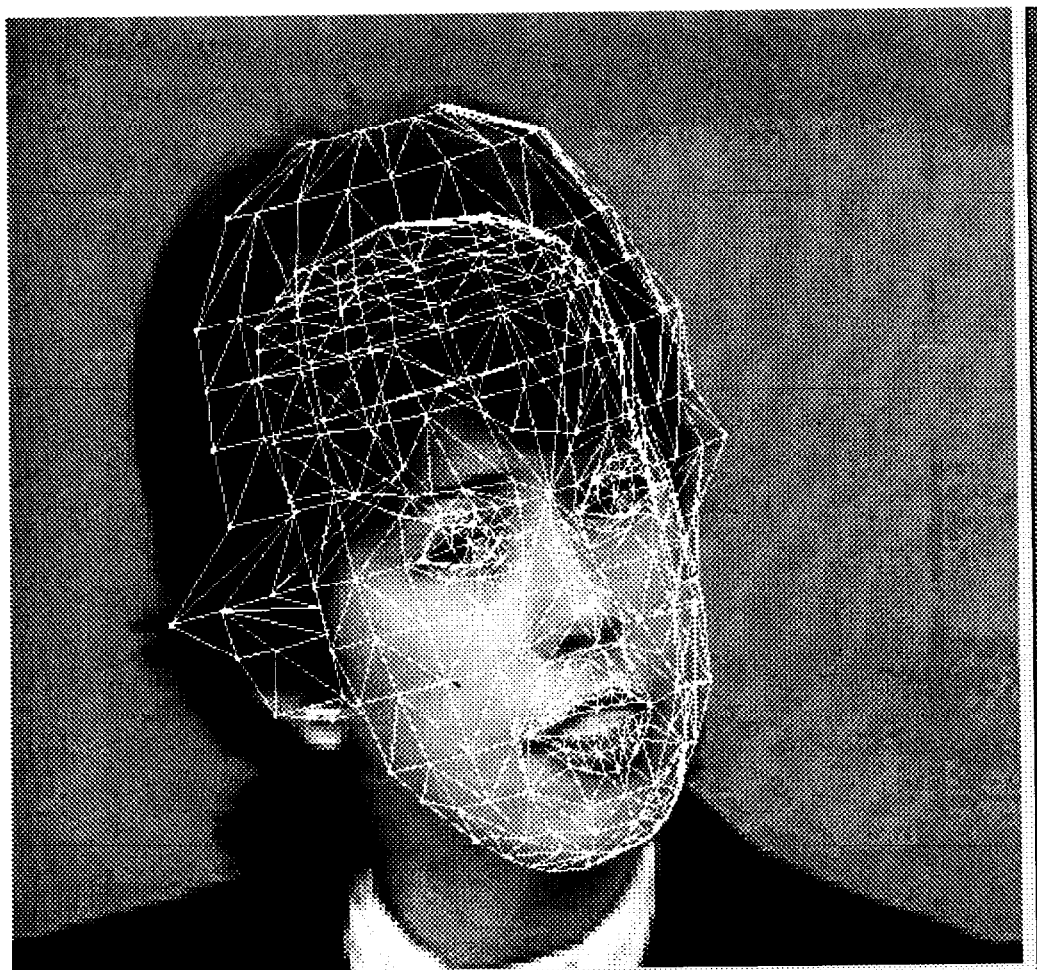


図 4.7(a) 4.3.1 の近似方式によるグローバル動きパラメータで合成したワイヤフレームモデル(CCITT Claire, 第 10 フレーム)



図 4.7(b) 4.3.1 の透視投影の検出方式によるグローバル動きパラメータで合成した画像(CCITT Claire, 第 94 フレーム)

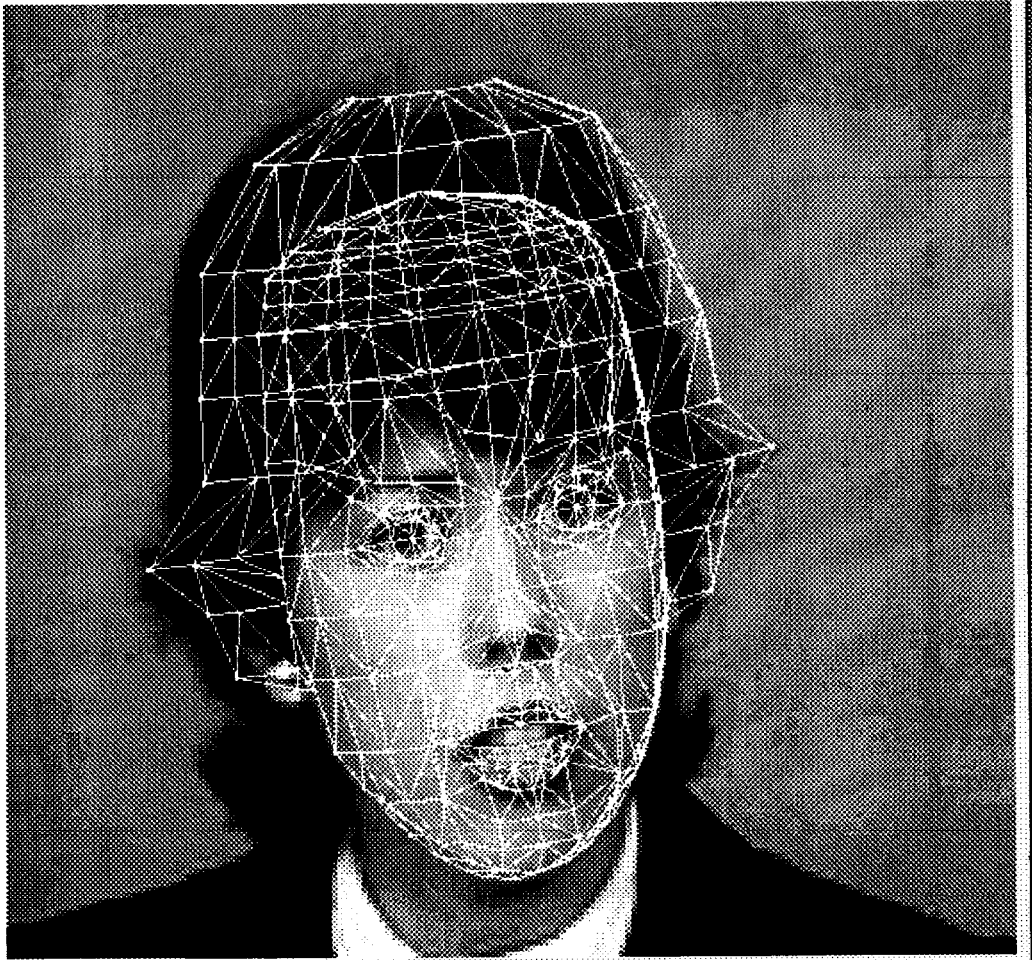


図 4.7(c) 4.3.1 の透視投影の検出方式によるグローバル動きパラメータで合成したワイヤフレームモデル(CCITT Claire, 第 112 フレーム)



図 4.7(d) 4.3.1 の透視投影の検出方式によるグローバル動きパラメータで合成した画像(CCITT Claire, 第 112 フレーム)

表 4.3 ビットレートの内訳 (Table : Detailed Items of Bitrate)

Item	Stationay State	With Initial Image
Input :ITU-T Claire, 4-118 fr 10 Hz (39 coded frames)		
Initial Image Information (QCIF 176x144,JEPG 1bit/pel)		176 x144 = 25344 bits
Initial Wireframe Fitting Points* (8x8) bit x 58 =	464 bits	→
Global Motion Parameters (8 bit x 8 = 64 bits / frame) 64 bit x 39 fr =	2496 bits	→
Local Motion Parameters Address for coded feature points 8 bits x 173 times =	1384 bits**	→
X,Y coordinate cooefficients (8+8) x 173 times =	2768 bits	→
Total	7576 bits	32920 bits
per frame	194.26 bit / frame	844.10 bit / frame
per second	1942.60 bit / second	8441.0 bit / second

*The initial fitting points are face contour 31 points, brows 8, eyes 10, nose 4, mouth 5

**The numbers of local motion points for coded frames indicated by an operator are shown in Table4.4

表4. 4 Table 2 The Numbers of Local Motion Points for Coded Frames

4fr	0	13	2	22	3	31	3	40	3	49	4	58	4	67	6
7	0	16	0	25	1	34	4	43	5	52	4	61	4	70	4
10	2	19	0	28	1	37	3	46	4	55	4	64	4	73	5

76	5	85	7	94	6	103	12	112	5
79	6	88	5	97	6	106	10	115	6
82	7	91	1	100	11	109	12	118	4

The left column is frame number and the right column is the number of coded local motion points in a pair of consecutive two columns .

第4章 グローバル動き検出とノイズ除去フィルタ

いる。後の章で、これに付加情報を加え、より汎用的で高画質化可能な符号化方式として実験を行なう。ローカル動きパラメータとしては、目の開口率と口の主な特徴点の座標を示すものを用いた。表4. 4に上記実験でのCCITTのClaireのローカル動きパラメータの符号化を行った座標値の個数の内訳を示す。ローカル動きパラメータは最大256種類の符号を想定し、8ビットで符号化している。また、初期画像はJPEG方式によって1ビット／画素の符号化がなされたものを用いている。

4.5 ノイズ除去フィルタと検出精度

4.5.1 インタラクティブ操作と入力ノイズ

知的符号化を映像ネットワークにおける応用に適用していくためには、画質を確保するために必要なインタラクティブ操作の内容を明確化することが必要である。人間が丹念に操作を行えば画質は向上して行くが、インタラクティブ知的符号化では操作者の技巧に依存するのではなく、一定の難易度の操作で符号化処理が遂行できるシステムである必要がある。本論文では、グローバル動きパラメータを検出するための特徴点入力における人間の操作において不確定な変動が加わることによる影響を調べ、これを取り除く距離拘束のフィルタを開発し、その有効性について述べる。次に、インタラクティブ操作による特徴点入力に関し、3次元ワイヤフレームモデルから入力された特徴点の奥行きを求める際に使用する最小2乗法の適用の仕方について検討する。すなわち、入力条件の個数は後に行うグローバル動きパラメータ検出における最小2乗法適用時の条件数に比べて少ないが、精度評価を行うことによって、条件数が妥当な範囲にあることを示す。更に、実験により本方式によるノイズ除去が有効に行われることを明らかにする。

図4.8に上で述べたインタラクティブ知的符号化システムの送信部に入力ノイズ除去フィルタを加えた場合の構成図を示す。符号化されるパラメータは初期画像、ワイヤフレームモデルのフィッティング情報、グローバル動きパラメータ、ローカル動きパラメータである。ローカル動きパラメータのうち目の形状はまばたきの度合いという一種の意味的なコマンドで表現されている。太枠部が本節で主に検討の対象になっている部分である。

顔画像を見ながら、12個の特徴点を入力するインタラクティブ操作のマンマシンインターフェースとして、マウスで画素位置を指定して入力できるツールを開発した。図4.9はグローバル動きパラメータ検出に関するツールの動作画面例である。まず、第1フレームに対するワイヤフレームモデルのフィッティングは終了しているものとする。入力画像系列を選択後、第2フレームより順次処理するとした場合、まず、初期画像である第1フレーム上にフィッティングされたワイヤフレームモデルを表示する。この時ワイヤフレームモデルの頂点間を結ぶ線分については全てを表示すると顔

画像が見えにくくなるため、本ツールでは間引いた状態で表示している。第2フレームの顔画像の12特徴点の位置は、第1フレームよりずれていればその特徴点のみマウスボタンにより新しい位置が入力される。特に入力しない場合は前フレームからの変化がなかったと見做し、前フレームの特徴点座標の位置が使用される。以下、次フレームの特徴点のうち前フレームからの変化のある特徴点のみを入力していく。全フレームに対する特徴点を入力後、グローバル動きパラメータ検出ボタンをクリックすることにより、グローバル動きパラメータ検出が実行される。

ここで、本グローバル動きパラメータは透視投影の設定の下で検出がなされるが、その関係は次のようになっている。図4.3に示すように、入力された2次元画像から想定される3次元物体面までの距離（約2～3m程度）を Z_{obj} とし、焦点面から距離 Z_{obj} の位置までの透視投影逆変換を行う。グローバル動き検出の第1段階の処理により、奥行き z を確定する。このとき x 、 y 座標は f を焦点距離として、実距離 $Z_{obj}+z$ により、

$$(Z_{obj}+z)/f \text{ 倍}$$

に修正される。次に、回転の検出は Z_{obj} を中心に行われる。 Z_{obj} は頭部ワイヤフレームモデルの中に固定された一つの中心に設定されている。受信側ではこれと同様の変換を行うが、ここで物体までの距離 Z_{obj} を変化させると再生画像の遠近感を変化させることができる。これにより再生画像において前後の動きを誇張したり、抑圧したりする効果をもたせることが可能である。本方式は別に報告するローカル動きパラメータ検出ツールと合わせ、システム化されている。

なお、図4.9に示した顔画像は512x480画素で表示され、12個の特徴点は精度を保つため直径1画素の円で表示して入力操作がされるが、本論文では印刷後の判読性を考慮し、直径5画素の円に拡大して示してある。

このようにして入力された2次元の12個の特徴点座標を用いて最小2乗法により3次元動きパラメータを求め、初期画像にフィッティングされたワイヤフレームモデルを変形することによって、動画像系列を生成する核となる部分が形成される。これに加えて主に表情などによる顔の各部品に応じた局所的変形を示すパラメータの検出がなされる。受信側では3次元ワイヤフレームモデルに初期画像の対応部分がテクスチャマッピングされ、動画像系列が再生される。

以上のようなインタラクティブ知的符号化における合成画像を観察したところ、時間的に変動する不自然な動きが見られた[59][60]。この変動の要因は特徴点の位置をマウスで入力する際の人間の入力操作に誤差が加わるためと考えられる。この誤差を

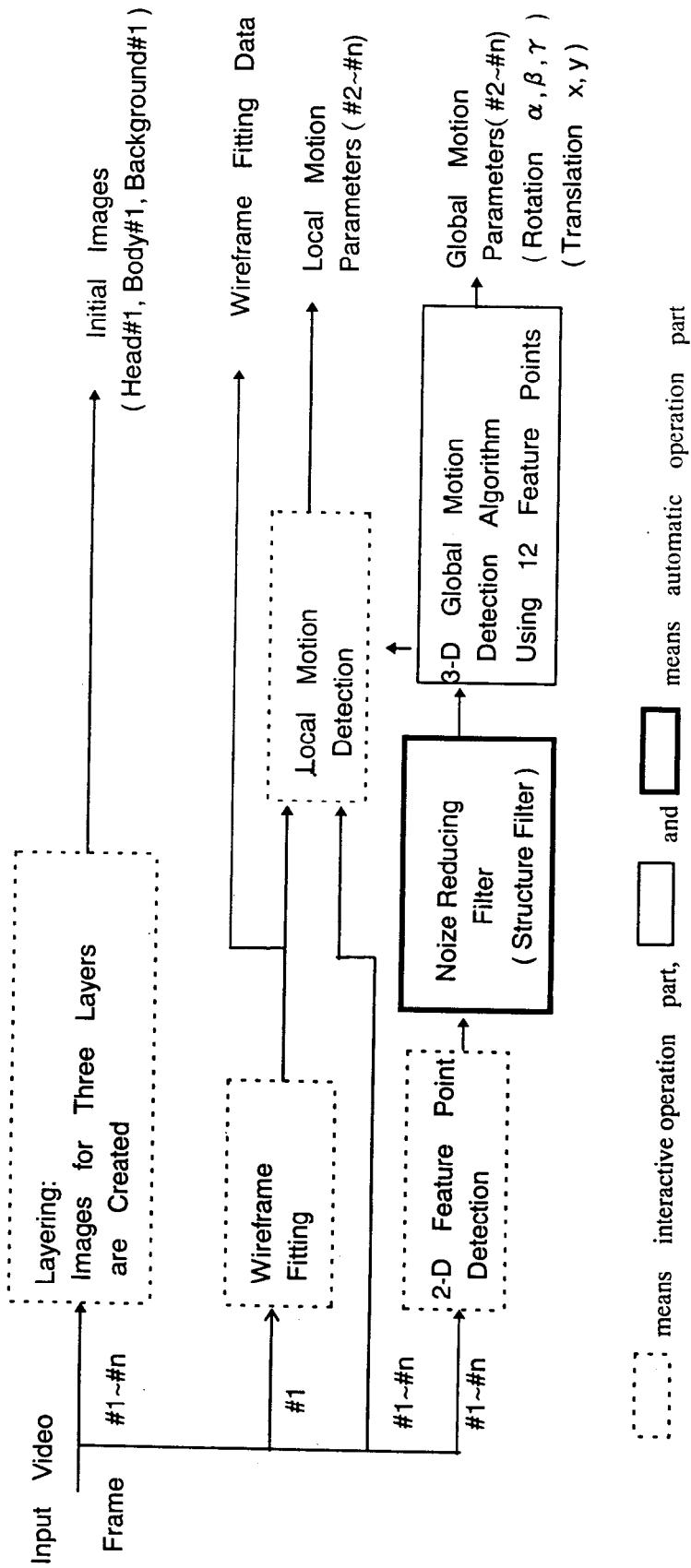


図4.8 ノイズ除去フィルタを加えたインタラクティブ知的符号化システムの符号化部
Encoder of interactive model-based coding with noise reducing filter.

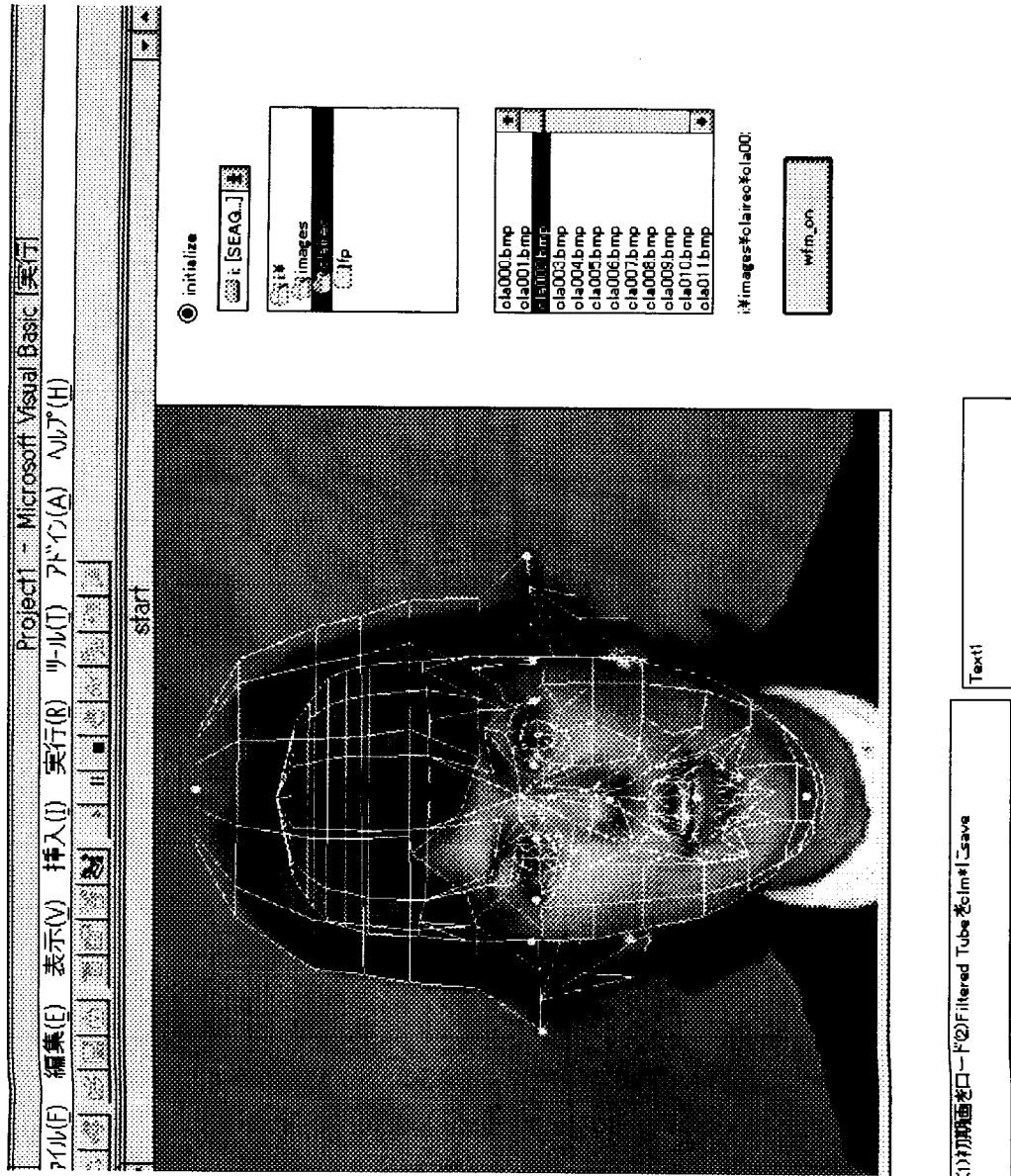


図 4.9 グローバル動きパラメータ検出を行うインタラクティブ操作ツールの動作画面例
 Displayed image in interactive operation tool for global motion parameter detection.

操作者に負担をかけることなく抑圧することが必要である。ところで、インタラクティブ操作においては操作量や操作の精度を規定しなければ、操作の過程が曖昧になる。入力操作の手順をマニュアル化でき、だれでもその仕様に従った入力を行えば、あとはシステムが対応する構成であることが望ましい。本論文では特に頭部のグローバル動きパラメータ検出のための入力操作の仕様化を、入力変動を除去するためのフィルタによって実現することを目指す。次に入力特徴点からグローバル動きパラメータを求める際に使用する最小2乗法の適用条件は、インタラクティブ知的符号化全体で、統一された基準であることが望ましい。そこで、最小2乗法の適用にあたり、精度の明確化をはかり、適用の妥当性を検証していくことを試みる。

4.5.2 最小2乗法の適用条件

はじめに2次元の12特徴点の座標値から、12個の特徴点の奥行き z_i と回転3個と平行移動2個のグローバル動きパラメータを求める際の、最小2乗法の適用の仕方について整理しておく。第1段階で、入力された12個の2次元特徴点座標 (X_i, Y_i) から奥行き z_i を求める。2個の特徴点から奥行き z_i を未知数とする1個の条件式が得られる[61]。12個の特徴点からは ${}_{12}C_2=66$ 個の条件式が得られる。未知数 z_i の数は12個である。このような設定で最小2乗法を適用し、 z_i を求める。次に第2段階では、12個の特徴点座標 (X_i, Y_i) と第1段階で求めた奥行き z_i を用い5個のグローバル動きパラメータに関する条件式が24個得られる。なお、グローバル動きパラメータの平行移動成分のうち奥行き z_i は最終的な描画結果が2次元表示されるため、不必要であるので求めない。この条件を用い最小2乗法によりグローバル動きパラメータを求める。最小2乗法の適用にあたっては、厳密にはまず次の3項目の前提が満たされていることが必要である[61]。

- ①測定値の誤差は不偏である（平均が0である）。
- ②測定値の分散は既知である。
- ③各測定値は互いに独立であり、共分散は0である。

しかし、通常これらの前提を検証することは不可能な場合も多く、本論文においても求めた結果の合成画像の形状が画面上で歪みが見えないことを以って許容する方針を採用する。また4章では未知数に対する条件の数を全体として統一されたものにするため、精度の評価を行っている。

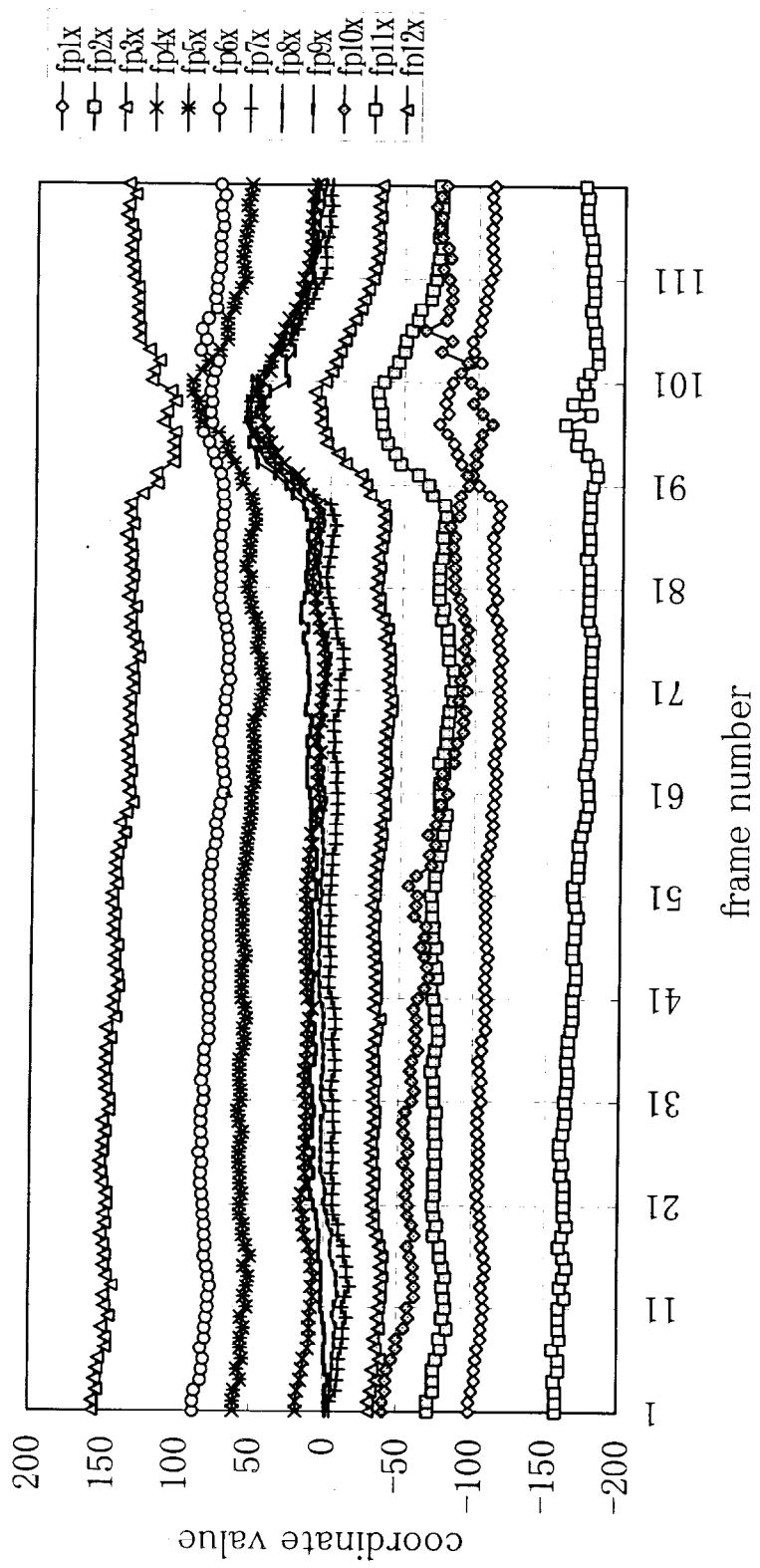


図 4.10(a) 特徴点のX座標の動き X-coordinate values of feature points.

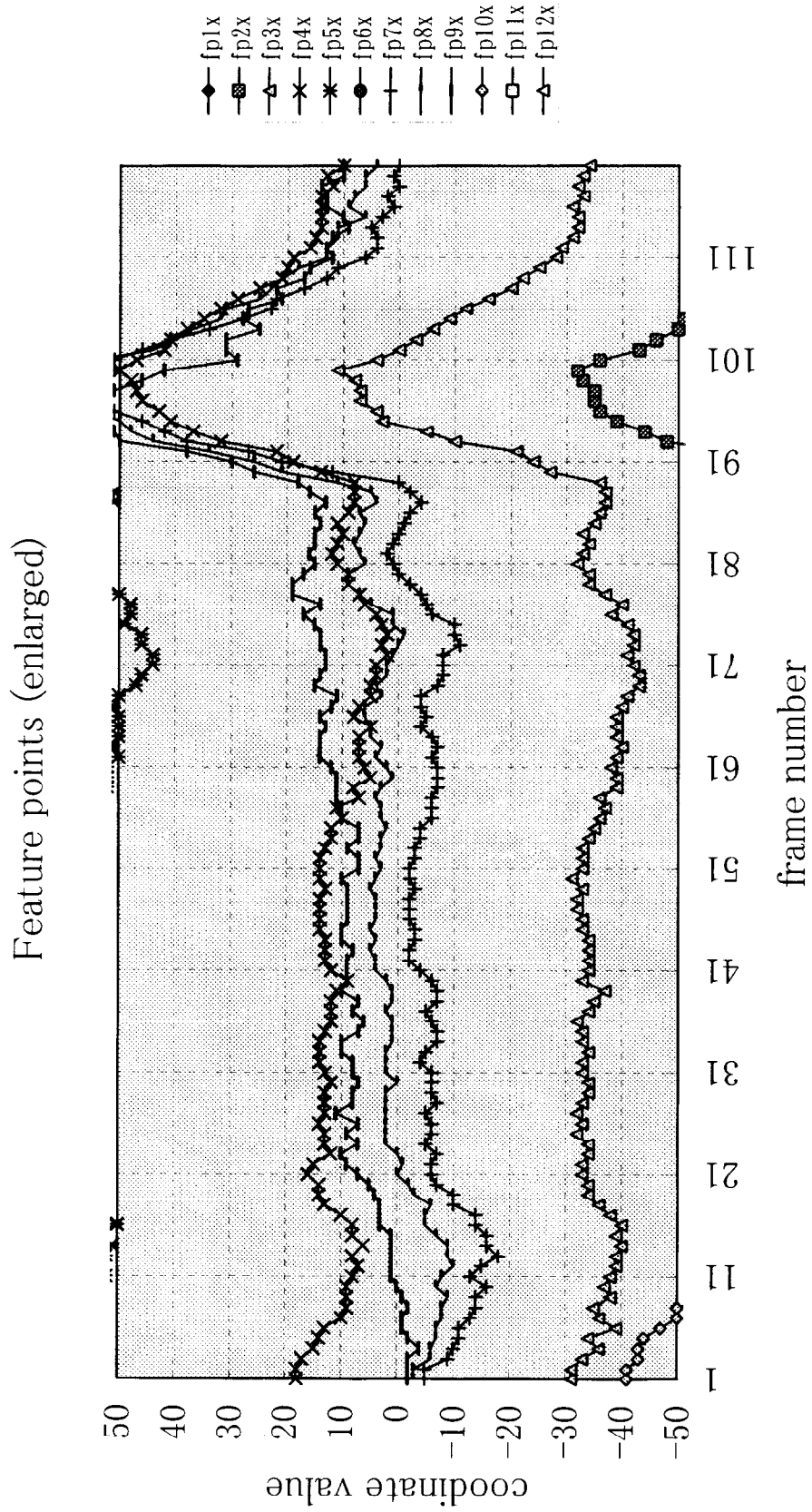


図 4.10(b) 特徴点のX座標の動き(拡大図) X-coordinate values of feature point(enlarged).

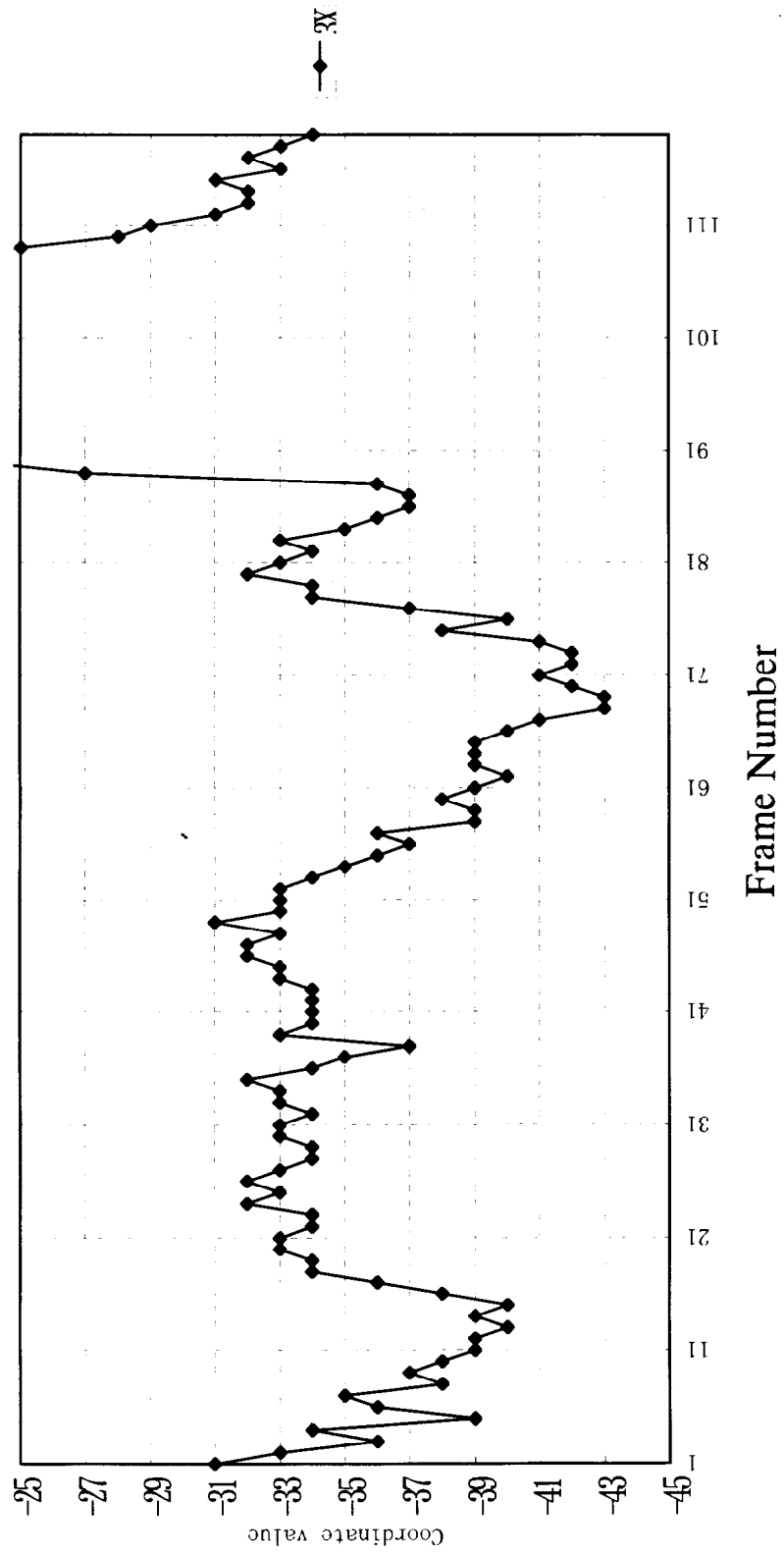


図 4.10(c) 特徴点"3"のX座標の動き X-coordinate values of feature point "3".

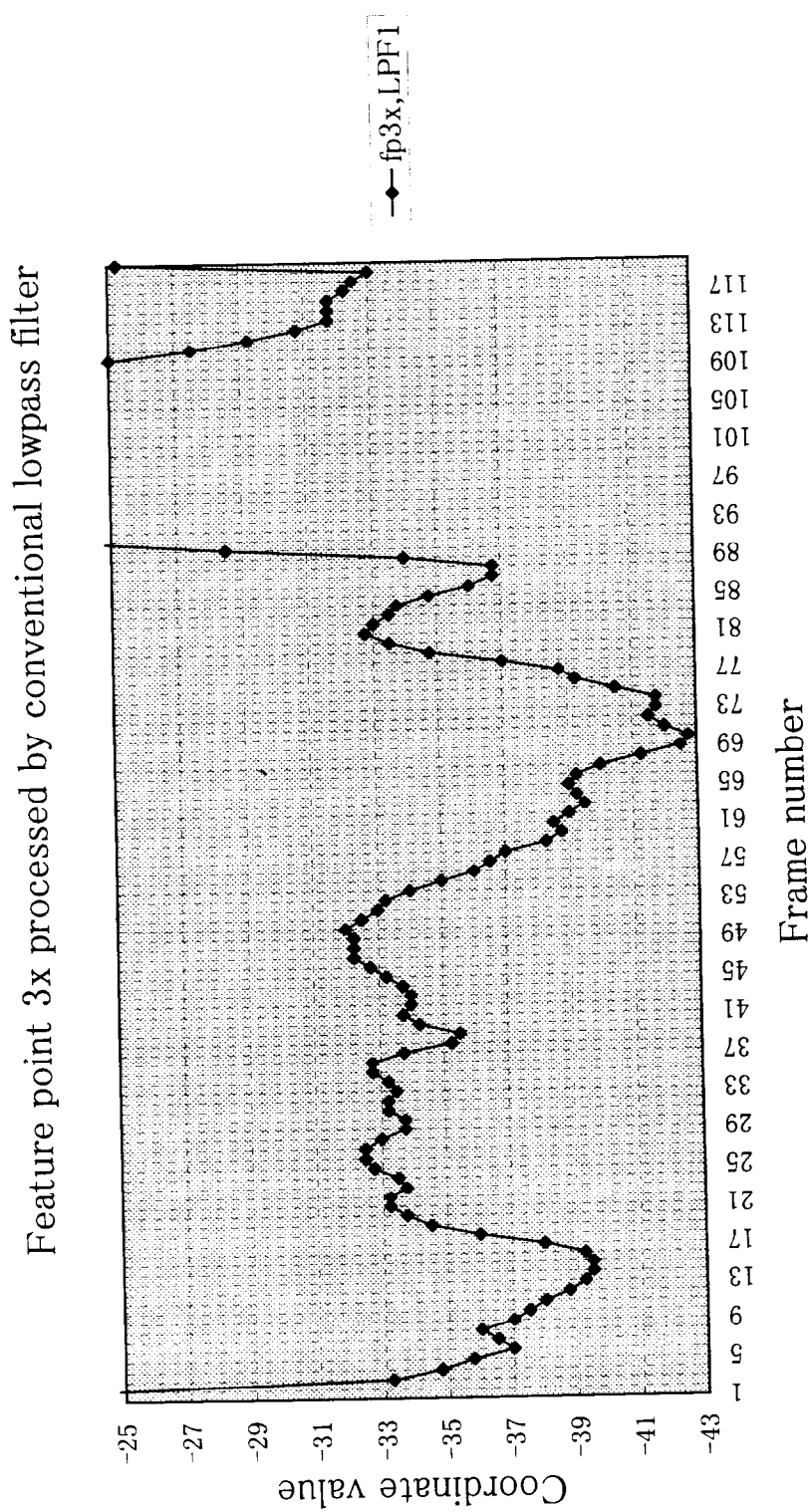


図 4.11 特徴点にLPFを施した後のX座標の動き X-coordinate values of feature point with LPF.

4.5.3 入力ノイズとLPF

上記のようにして求めたグローバル動きパラメータを用いて画像を合成したところ、不自然な振動的動きが観察された。この様子はワイヤフレームモデルのみの系列の合成画像でも観察された。図 4.4 は使用した 12 個の特徴点の具体的な配置図である。次に、図 4.10(a)は CCITT の標準画像 Claire に対する 12 個の特徴点の座標値をプロットしたものである。図 4.10(b)に一例として特徴点 3 の x 座標値のみを取り出して拡大したものを示すが、振動的な変化をしている部分があることが分かる。このような振動的な動きに対し、まず時間方向に対する LPF の適用を検討した。図 4.11 は 3 タップ(0.25,0.5,0.25)のフィルタを時間方向に適用した結果であり、信号としては振動が抑圧され一見問題は解決したかのように思える。しかし、合成画像を動画像系列として表示してみたところ、依然として不自然な動きが観察された。

この不自然な動きの原因としては次のようなことが考えられる。一つは時間方向にフィルタを施すことにより各特徴点ごとに独立に振動は抑圧されるが、独立でない成分は除かれないことである。また、顔画像のワイヤフレームモデルに対して、グローバル動きパラメータ検出においては剛体仮定をとっているが、実際は非剛体の変形があるため、特徴点入力を正確に行ったとしても変形を引き起こす可能性もある。また、顔画像を対象としているため、信号レベルとしてはある程度抑圧されても、顔面上ではたとえ 1 画素レベルの変化であっても、それによる形状や表情の変化は人間の目による観察では敏感な影響として現れることも考えられる。そこで、各特徴点を独立に扱うのではなく、相互の距離を拘束する作用を持つフィルタ、ないしはノイズに対して選択的に作用し、フィルタの減衰量を上げてても信号成分が維持されるような新しいフィルタを構成する必要がある。

まず、構成上線形な演算のみに限定する必要はないので、従来よりある非線形フィルタについての調査を行った。画像信号処理では rank-order に関する非線形フィルタについて Coyle、Fong のサーベイがある[62][63]。Coyle によればメジアンフィルタ、スタックフィルタ、順序統計フィルタ、モルフォロジカルフィルタなどのクラスがあり、それらに巡回型、一般型、多段型、直列型などの各変形例がある。これらはエッジの保存や再生、ノイズ除去に有効である。中でもメジアンフィルタは簡単な構成でノイズ除去を有効に行えるが、ウィンドウ内の rank-order 情報のみしか使用していないため、時間的關係が失われている[64][65]。また、NIPS95 ではいくつかの非線形フィルタが提案されている問題があり、ジッタやストリーキングの発生が指摘さ[66]。以上の中で、ベクトルメジアンフィルタ[67][68]は本論文で必要となるフィル

タに類似した入力条件を備えている。これは、テレビジョンのカラー3信号や衛星マルチスペクトル信号処理において、複数の信号をベクトルとして扱うものである。ベクトルメジアンフィルタはベクトルのノルムを最小化する基準でノイズ除去を行うものであるが、一括処理という点では考え方の基になるものである。本論文では、複数の顔画像上の特徴点を全て独立に扱うのではなく、顔の形状と動きを形成する点として相互に関係を保ったまま、処理を行おうとしている。

4.5.4 仮想不動点の作用

本論文で提案するノイズ除去フィルタの構成について述べる。3次元空間内では特徴点間の距離は不変である。一方、2次元画像面へ射影された特徴点間を結ぶ線分の長さは変化し得る。3次元空間座標 (X,Y,Z) に対し、2次元焦点面座標 (x,y,z) は、

$$x=X*z/F, \quad y=Y*z/F \quad (F \text{ は焦点距離})$$

の関係がある。今、3次元空間内での2点 P_1 、 P_2 間の距離を L 、ベクトル P_1P_2 と x - y 平面のなす角を θ とすると、 L の焦点面への射影は $L'=L*\cos \theta$ となる。回転に対し L' は $\cos \theta$ に比例し、連続でその微分係数は ± 1 の範囲内にある。従って2次元平面上の距離 L' の単位時間あたりの変動には上限がある。この変動に拘束をかけてもデータの変化は妥当な範囲に収まると考えられる。ローカル動きは非剛体の動きを含むので、上記の剛体仮定の変形量に加算されることになる。またグローバル動きパラメータ検出の間だけ一時的に3次元ワイヤフレームモデルを剛体と仮定し、3次元モデルに基づいた回転を行ったとしてもローカル動きもあるため、ノイズ除去が正しくなされとは限らない。従って、理想的な3次元剛体回転の条件を維持すると、ローカル動きによって大きい破綻を来す恐れもある。そこで、入力された2次元平面上での特徴点の座標に対し、その距離を安定化させる手法を検討する。

まず、2点間距離に関してのノイズを除去するフィルタを考える。この場合、2点を結ぶ方向に対するノイズの成分は除去可能であるが、それと直交する方向の成分には効果がない。極端な例として、2点間の中点を中心として回転するノイズに対しては全く抑圧効果がない。また変形を伴っている特徴点座標間距離を相対的な位置関係のまま取り扱っても、上記ノイズを的確に除去することはできない。

このようなことはフィルタを施すにあたり基準とする点が存在しないために、処理後の点の位置が不確定になるためと考えられる。そこで処理を施す顔画像上に仮想的な不動点を設け、その不動点を基準として長さに関する処理を行うことが有効と考えられる。各入力フレームの顔画像の中心に仮想的な不動点を設けると、この理想的

第4章 グローバル動き検出とノイズ除去フィルタ

な点はフレームごとに異なった位置をとる。この仮想的な不動点を用いたフィルタは顔の構造情報を生かした処理を行うものであり、その処理は非線形演算になって来る。このフィルタそのものの構成を定義する前に、特徴点間の距離の時間系列に対する線形な帯域通過フィルタ (LPF) の適用を考えておく。

x_1, x_2 を2つの特徴点の x 座標として以下説明する。差分 $x_3 = x_1 - x_2$ は線形演算順序に対し可換である。すなわち、 $h(n)$ をフレーム番号 n (時間方向) に関するフィルタの長さ $2L+1$ のインパルス応答とすれば、 $x_1(n), x_2(n)$ に対するフィルタ出力は

$$u_1(n) = \sum_{k=-L}^L x_1(n)h(k-n)$$

$$u_2(n) = \sum_{k=-L}^L x_2(n)h(k-n)$$

となる。 $x_3(n) = x_1(n) - x_2(n)$ とすると、 $x_3(n)$ の出力は、

$$u_3(n) = \sum_{k=-L}^L x_3(n)h(k-n) = u_1(n) - u_2(n)$$

となる。このように差分 $x_3(n)$ に対する線形フィルタは各点に独立に施すフィルタの差分に分解され、基準の位置からの長さというものには直接的な拘束を与えていない。そこで、 $x_2(n)$ を顔画像上に仮想的に定めた不動点とすると、この点を基準に長さに関する拘束を与える処理を有効に行うことができるようになる。 $x_2(n)$ を不動点とした時、 $x_3'(n) = x_1(n) - x_2(n)$ に上記 LPF を適用すると、

$$u_3'(n) = \sum_{k=-L}^L x_3'(n)h(k-n) \quad (4.19)$$

となり、 $x_1(n)$ は

$$x_1(n)' = x_2(n) + u_3(n)'$$

と表される。このように不動となる基準点を設け、長さに対する LPF を適用することが本論文における提案方式の特徴となっている。

ここで仮定している不動点としては、座標の原点のように画像と独立した点では

なく、顔画像中の意味のある点、例えばその人の鼻の頭など被写体である顔の中では同一の点を想定している。従って、撮像後の画像としての座標値は変動しうるものである。今、このような仮想的な不動点の代わりとして特徴点の一つを割り当てることを考える。選んだ特徴点が不動点として信頼できるものであれば良いが、特徴点はインタラクティブ操作により入力されたもので、どの点もある程度までしか信頼できない。そこで不動点としての役割を互いに順次交代し、各特徴点に持たせるような構成を考える。

4.5.5 構造型ノイズ除去フィルタ

このような不動点に基づくフィルタの動作の様子を模式的に示したものが図 4.12 (a-i)である。(a-c)は直流信号に同相の振動的ノイズ成分が加わった場合、(d-i)は直流信号に逆相の振動的ノイズ成分が加わった場合の例である。原信号(a)に対し前記3タップのLPFを1回施した場合が(b)、2回施した場合が(c)である。また片方の点、例えば x_2 が不動点の役割になったとき、このような同相の変動は除去されないで、図 4.12(a)の形が保たれる。一方、逆相の場合、(d)の原信号に対し前記のLPFを1回施した場合が(e)、2回施した場合が(f)である。 x_2 を不動点と仮定した場合の同一のLPFを x_1 に施した出力が(g)、この(g)に x_1 を不動点として x_2 にフィルタ処理した出力が(h)、更に(h)に x_2 を不動点として x_1 にフィルタ処理を施した出力が(i)となっている。この例でわかるように、仮想不動点の役割を交互に入れ替えその距離の時間的変動に対しLPFを施すことにより、距離の局所的変動が雑音と見做され抑圧される。顔のワイヤフレームモデルの構造に着目し、仮想不動点を基にそこからの距離に拘束をかける処理を顔画像における構造型フィルタ(以下SFと略す)として提案する。具体的には(4.19)式で定義される処理を基本とし、不動点としては特徴点を順次採用して処理を行う。このSFでは1回ごとの処理で1点を固定しているため、2点間の距離方向のノイズ成分だけでなく前に述べたようなそれと直交する成分や2点の midpoint に関し回転する動きも抑圧される。一方、同相的な動きについては顔全体が同じ方向に動いていると見做すことができるので、同相成分が抑圧されないことは弊害にはならないと考えられる。単なる時間方向のLPFの場合は信号成分の一般的変動や急な動きなどに対し一様に影響があるが、本フィルタでは同相的な動きであれば

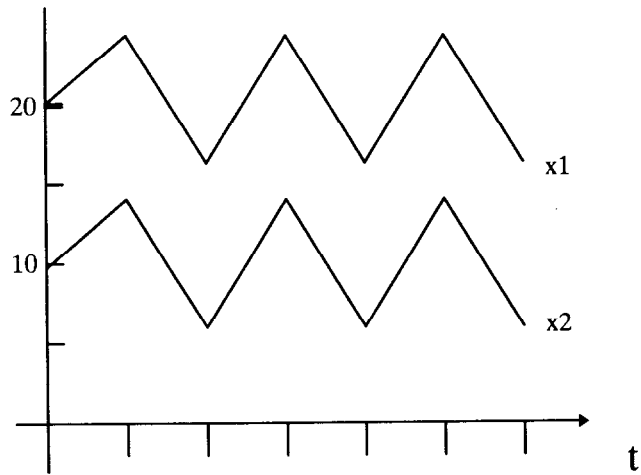


図 4.12(a) 同相ノイズ,未処理
In-phase, without filtering.

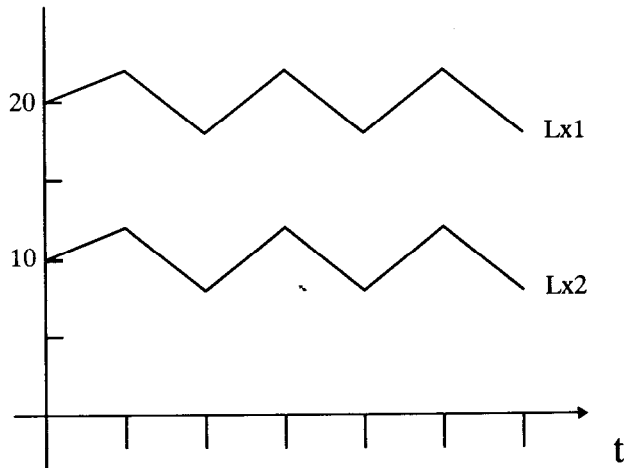


図 4.12(b) 同相ノイズ, LPF を1回適用
In-phase, with a single LPF operation.

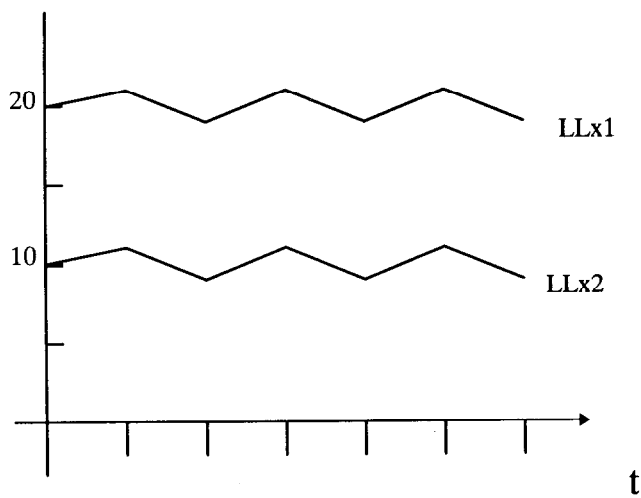


図 4.12(c) 同相ノイズ, LPF を2回適用
In-phase, with a double LPF operation.

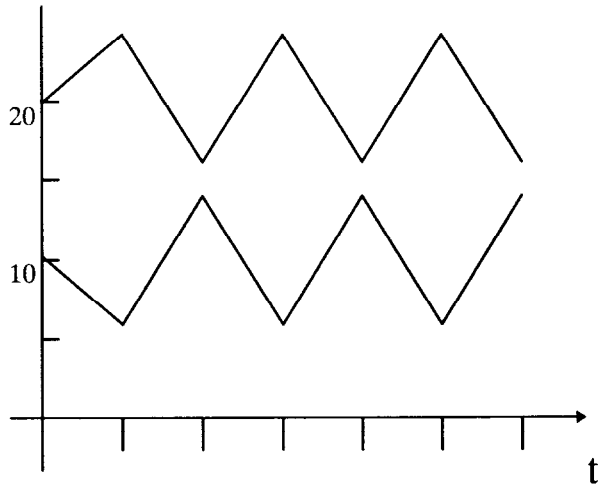


図 4.12(d) 逆相ノイズ, 未処理
Opposite-phase, without filtering.

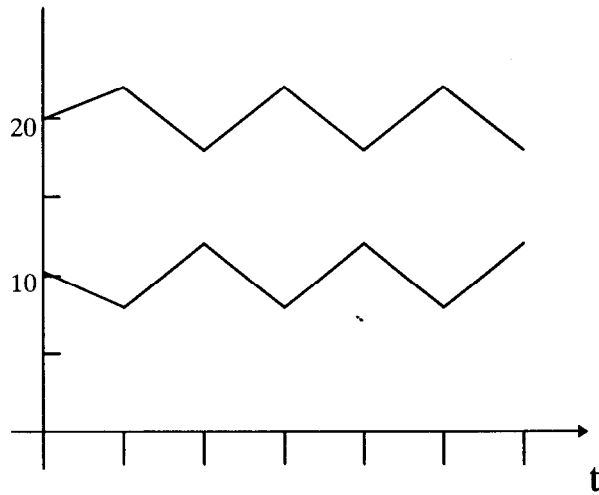


図 4.12(e) 逆相ノイズ, LPF を 1 回適用
Opposite-phase, with a single LPF operation.

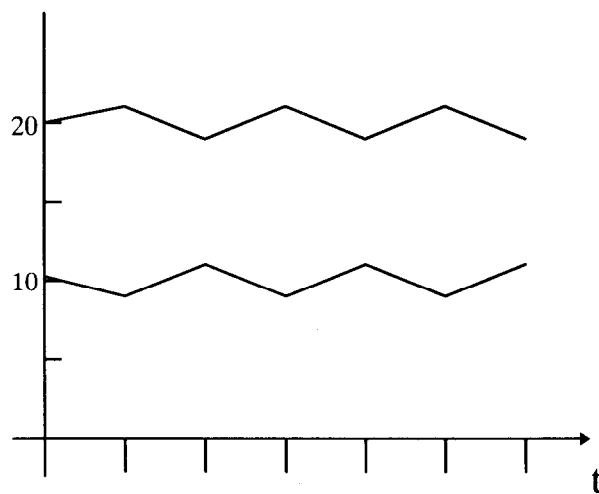


図 4.12(f) 逆相ノイズ, LPF を 2 回適用
Opposite-phase, with a double LPF operation.

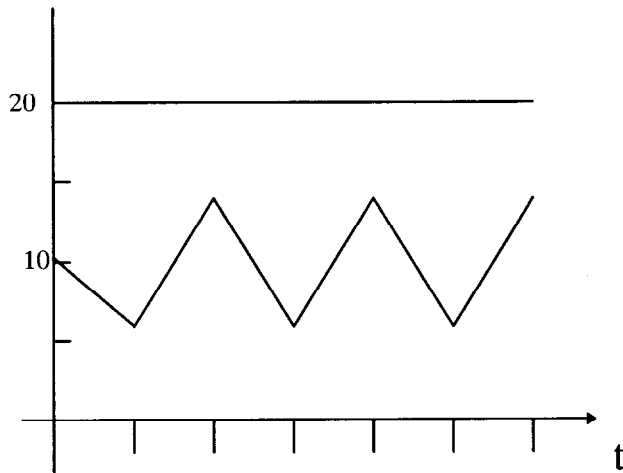


図 4.12(g) 逆相ノイズ, 提案方式のフィルタを1回適用
Opposite-phase, with a single filtering of proposed filter.

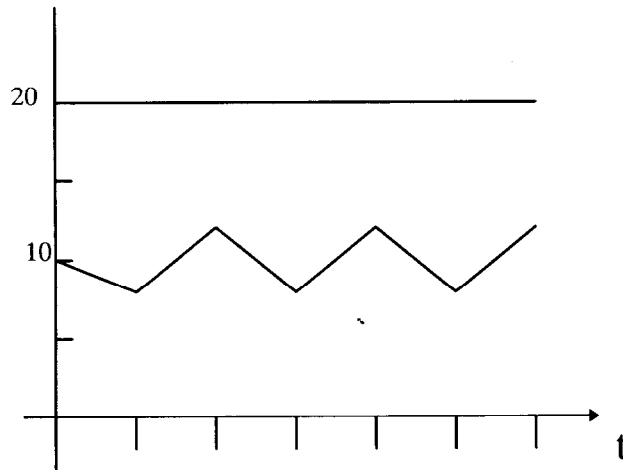


図 4.12(h) 逆相ノイズ, 提案方式のフィルタを2回適用
Opposite-phase, with a double filtering operation of proposed filter.

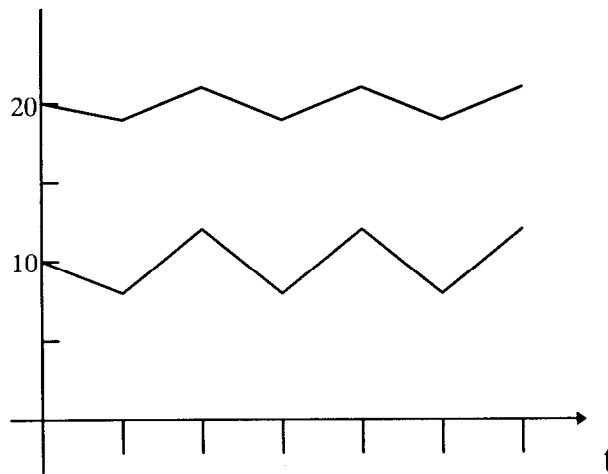


図 4.12(i) 逆相ノイズ, 提案方式のフィルタを3回適用
Opposite-phase, with a triple filtering operation of proposed filter.

全く影響を受けないというメリットがある。また、入力ノイズやローカル動きによる変動を全て含んだ特徴点座標データに対し、距離変動の抑圧がなされ全体の形状が整形される効果もある。以上、本章で提案したフィルタはワイヤフレームモデルで合成される顔画像の構造的な特性を利用している。顔画像処理においては、相対的な変動はわずかであっても再生後の画像においては敏感に検知される場合があるという特性を考慮すると、相互の距離関係を拘束するこのような構造型フィルタが有効であると考えられる。

4.6 特徴点の奥行き z の検出における精度評価

入力操作などで生じたノイズ変動成分を抑圧した特徴点座標値をもとにグローバル動きパラメータの検出を行う。グローバル動きパラメータ検出は4.3で述べたように2段階に分けてなされる。まず第1段階においては、特徴点を N 個とすると条件式が $N C_2$ 個になる。求める未知数は各特徴点の移動後の奥行き z であるので N 個である。最小2乗法適用における条件式の数はDixonのガイドラインによると、未知数の2乗個程度あることが良いとされている[55]。奥行き z を求める場合に文献[42]のアルゴリズムを用いるとき、未知数の2乗と条件式の数の比は

$$N^2 / N C_2 = 2N / (N-1)$$

で、図4.13(a)、(b)に示すようになっている。文献[48]では $N=12$ 、すなわち $N^2 / N C_2 = 2.18$ である。 $N \rightarrow \infty$ としてもその極限は2であり、精度の面では $N=12$ と大差ない。特徴点をマニュアルで与えることから、現実面を考慮して、 $N=12$ 程度が妥当であると考えられる。

次に、 $N \rightarrow \infty$ での精度について検討する。まず、最小2乗法を用いるにあたって設定された仮定①~③が成り立っているものとする。ガウス・マルコフの定理により最小2乗法による推定値は最良線形不偏推定量になっている[69]。すなわち、この推定量の分散はすべての推定量の中で最小となっている。このとき残差の平方和の平均 S_E は

自由度 ϕ の χ^2 分布に従う[69][70]。すなわち、

$$E(S_E / \phi) = \sigma^2$$

ただし、 σ^2 は推定値の誤差の分散、 E は期待値、

$$\phi = p - n, \quad p = \text{条件式の数、}$$

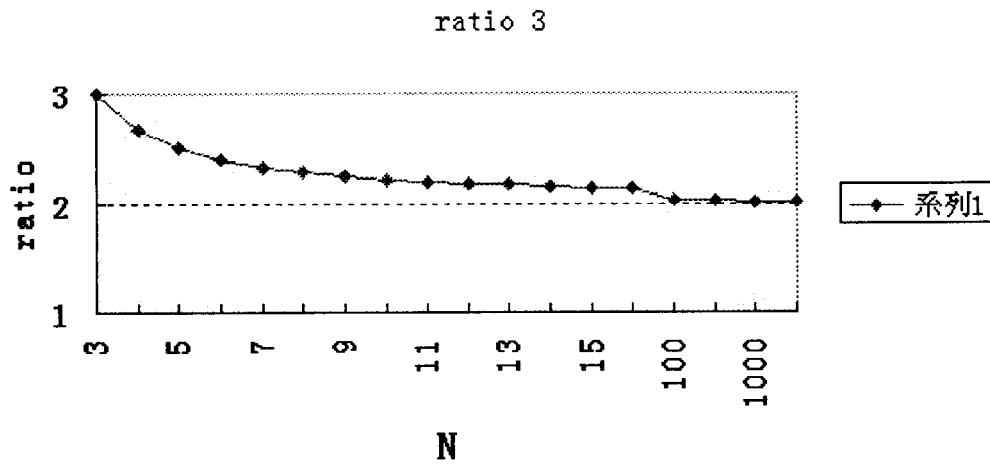


図 4.13 (a) 未知数の 2 乗と条件式の数の比

Ratio of the squared number of unknowns and the number of conditions.

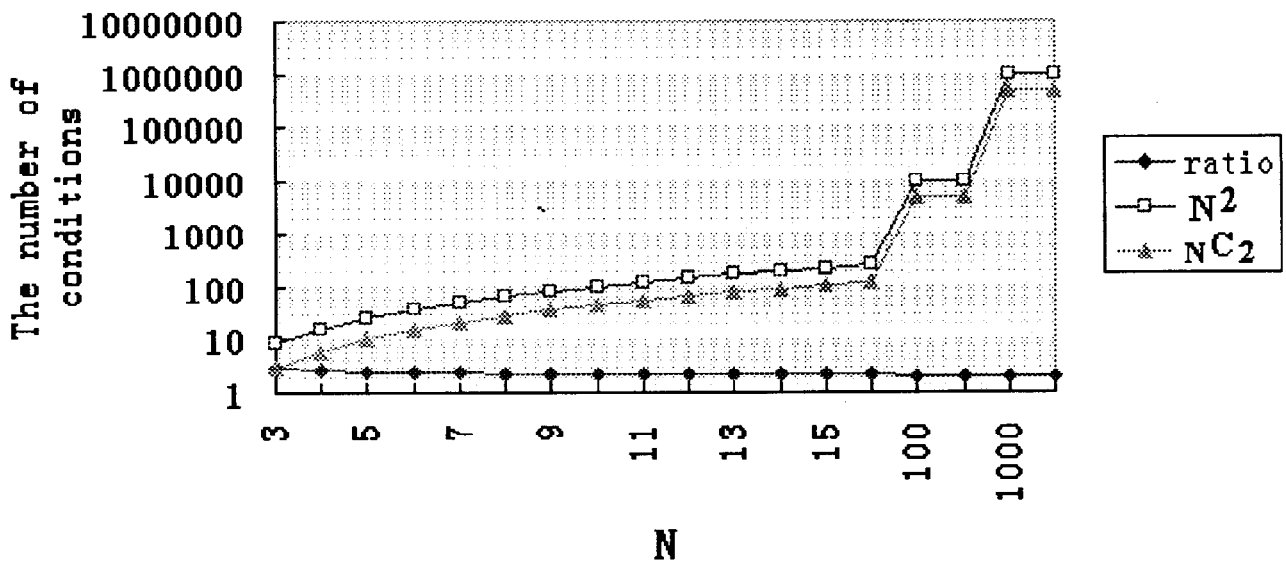


図 4.13 (b) 未知数の 2 乗と条件式の数の関係

Relation between the squared number of unknowns and the number of conditions.

第4章 グローバル動き検出とノイズ除去フィルタ

n =未知パラメータの数、

となる。これを上記の場合に適用すると $p=66$ 、 $n=12$ となり自由度 $\phi_1=54$ となる。一方 Dixon のガイドラインに従うとした場合、 $p=n^2$ となり、その場合の χ^2 分布の自由度は $\phi_2=144-12=132$ となる。

χ^2 検定法において、有意水準 $\alpha=0.01$ とした場合、残差の平方和の平均が $E(S_E/\phi)=\sigma^2$ であるという仮説を立てる。このとき χ^2 分布表より [71]、この仮説を棄却する領域の分散を求める。 $\alpha=0.01$ の χ^2 分布の χ^2 の実現値(推定値の誤差の分散)は、表 1 に引用してあるように本方式の場合、

$$\sigma_p^2=81/54=1.50$$

Dixon のガイドラインに従う時

$$\sigma_i^2=173/132=1.31$$

となっている。従ってその比は

$$\sigma_p^2/\sigma_i^2=1.15$$

である。これから本方式による推定値の誤差の分散は Dixon のガイドラインを実現した場合に対し 15%程度しか大きくなることがわかる。以上より、条件式の数 $p=66$ は妥当なものと思ふことができる。

4.7 実験と考察

前節までに述べた方式に基づき入力特徴点からノイズ成分の除去を行い、グローバル動きパラメータを求めた。4.5.5 で述べたフィルタの仮想不動点として、特徴点 3 から 8 を順次用いた。特徴点の座標値の時間的変動を図 4.14(a)~(h)に示す。図 4.14(a)に従来の LPF を 0~6 回施した時の変化を示す。図 4.14(b)に本論文で提案した構造型フィルタ(SF)を 0~6 回施した時の変化を示す。従来の LPF では回数が増加するに従い座標値の時間的変動が低域成分のみになっていくのに対し、提案 SF では回数による変化は少なく、高域成分も保存されている。図 4.14(c)にフィルタの無い場合、LPF(6 回処理)を適用した場合、及び本論文で提案する構造的フィルタ SF(6 回処理)を適用した場合の比較を示す。この比較図より LPF 法に対し SF 法においては、時間方向に不必要な平滑化がなされないという特性が期待できる。すなわち、LPF 法は図 4.12(a)~(I)で調べたようにノイズ成分と信号成分を同時に除去する欠点があるが、SF 法ではノイズ成分の除去を中心に作用する効果を持つ。

第4章 グローバル動き検出とノイズ除去フィルタ

図 4.14(d)、(e)は各々特徴点 3、4 に対する、フィルタ処理前後の座標値の時間的変動の一部を拡大表示したものである。図 4.14(f)は特徴点 3、4 間の距離に対するフィルタ処理前後の座標値の時間的変動である。フィルタ処理前に比べ、特徴点 3、4 間の距離は二つのフィルタでほぼ同じになっている。

図 4.14(g)に主な特徴点間の距離のフィルタ適用前を、図 4.14(h)に特徴点 3 と 4 の x 座標の距離のフィルタ適用後の様子を示す。特徴点間の距離については、LPF と提案 SF はともに同一の収束点に到達していることがわかる。これから、第一に特徴点間の距離の時間的変動を抑圧するため、形状保持の効果があること、第二に個々の位置は入力時の近くにあり、LPF 法の如く周波数特性を劣化させることが無い。このように提案 SF は雑音成分を除去する過程で、特徴点の位置を保存しながら相互の距離の変動を抑圧する効果があることが確認できる。又図 14(e)において第 39、40 フレームで逆転するように見える現象は、形状を保持する整形作用が働いたために生じたものと考えられる。LPF 後の形がほぼ相似であることから、特徴点 3、4 は距離を保ったまま並進していると解釈される。

図 4.15(a)、(b)は SF (6 回処理)を適用する前後の検出された回転パラメータである。(a)ではインタラクティブ入力操作のノイズ成分がパラメータ検出にも影響しているが、(b)ではノイズが除去されていることが分かる。

検出されたグローバル動きパラメータを基にローカル動きパラメータを求め再合成した画像においては、不自然な振動を示すようなノイズは除去されていた。主観評価実験によれば、再合成画像は従来の波形符号化方式と同等ないしはそれ以上の画質を示している[48]。この主観評価実験では CCITT の Claire、Miss America を用いた。波形符号化としては H.263 方式を用い、量子化切替で、約 40~180kbps まで 3 段階の復号画を比較の基準とした。提案方式はレート 54kbps、46kbps での符号化で、波形符号化で 80kbps と 60kbps 相当の評価が得られている。

図 4.14(e)の第 40 フレームで、最大 3 画素の整形がなされている。図 4.14(h)では距離が数画素にわたり変動している波形が処理されている。これから、インタラクティブ操作の仕様として、画素位置で 0~3 画素までの許容誤差範囲で人間が操作すれば、安定にグローバル動きパラメータが検出できることが分かる。この特徴点入力操作は特徴点位置をマウスで指示入力していくことで誰でも行えることであり、操作量、難易度ともに仕様として規定可能なものであるといえる。

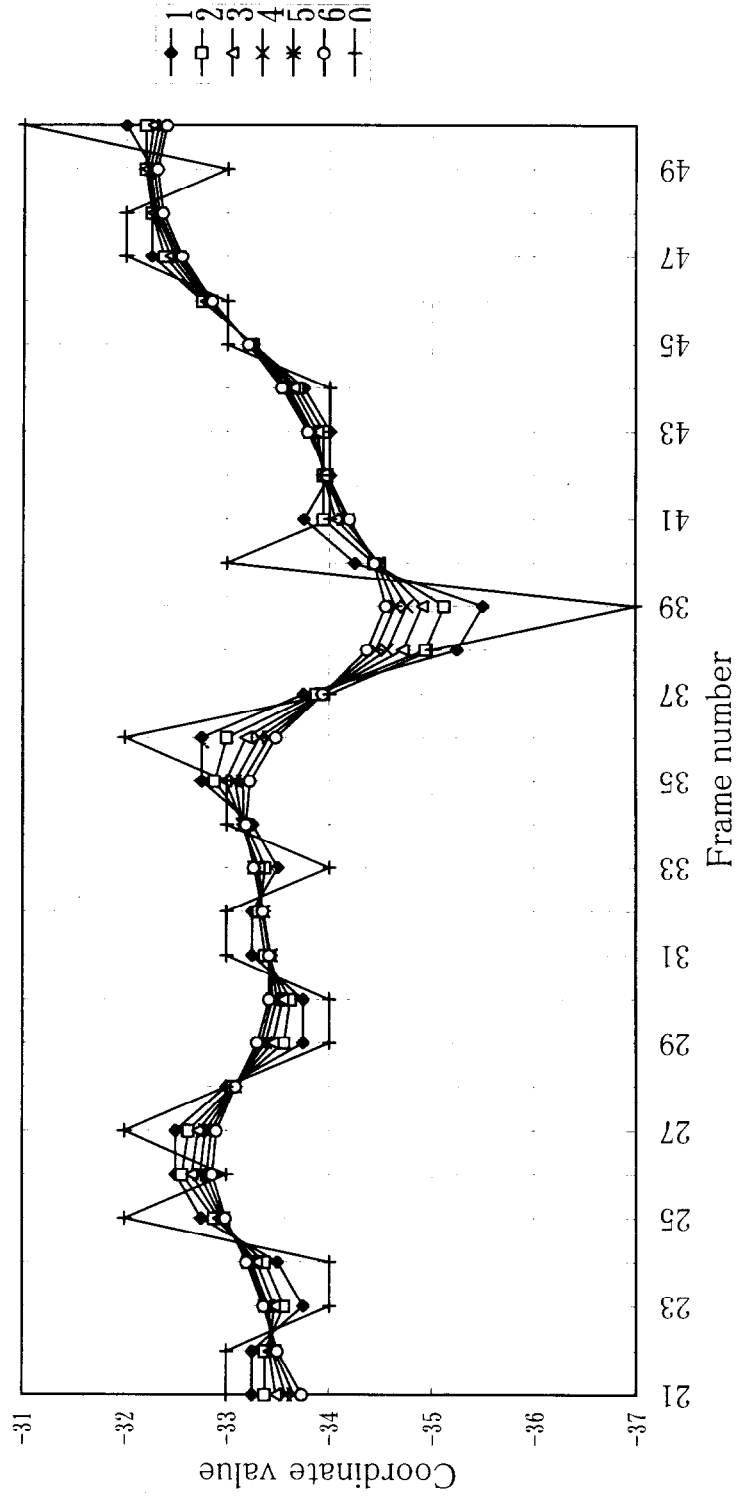


図 4.14(a) 特徴点"3"のX座標値にLPFを0から6回施した時の変化(縦軸の単位は画素間隔)
 Varies for X-coordinate values of feature point "3" operated by LPF for 0-6 times

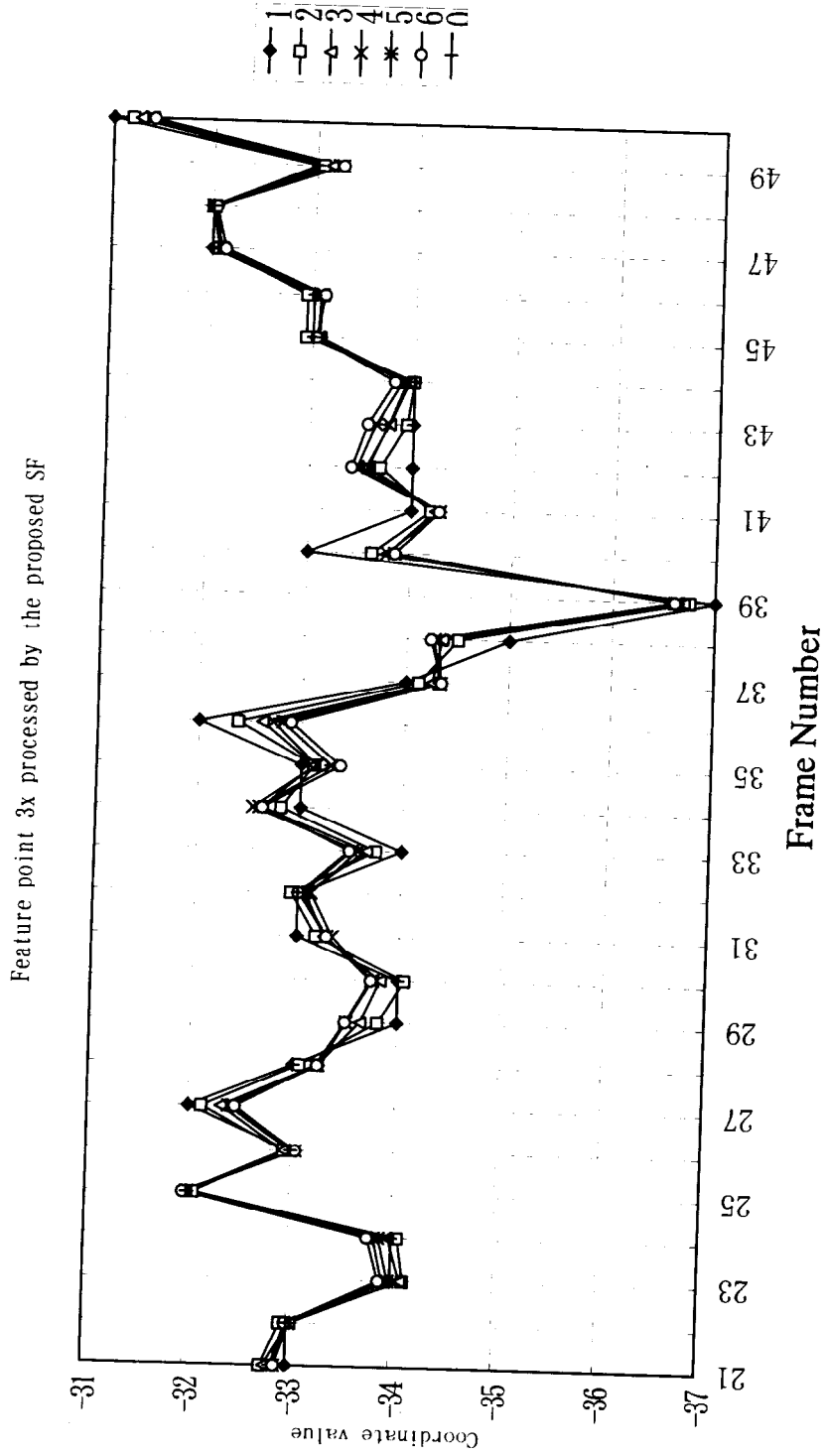


図 4.14(b) 特徴点"3"のX座標値にSFを0から6回施した時の変化 (縦軸の単位は画素間隔)
 Varies for X-coordinate values of feature point "3" operated by SF for 0-6 times

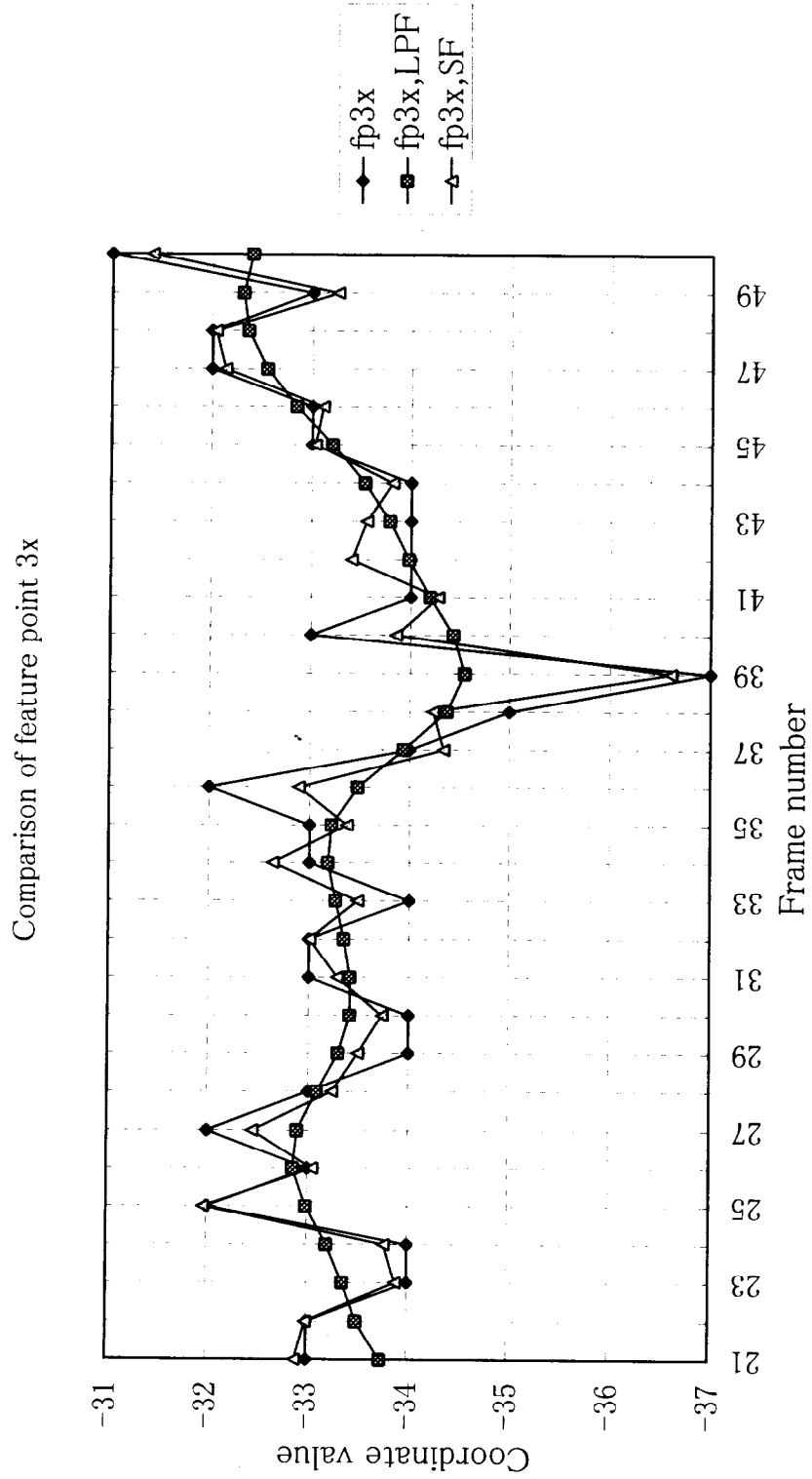


図 4.14(c) 特徴点“3”のX座標値にLPF,SFを6回施した時の比較 (縦軸の単位は画素間隔)
 Comparison between X-coordinate values of feature point “3” operated by LPF and SF after 6 times

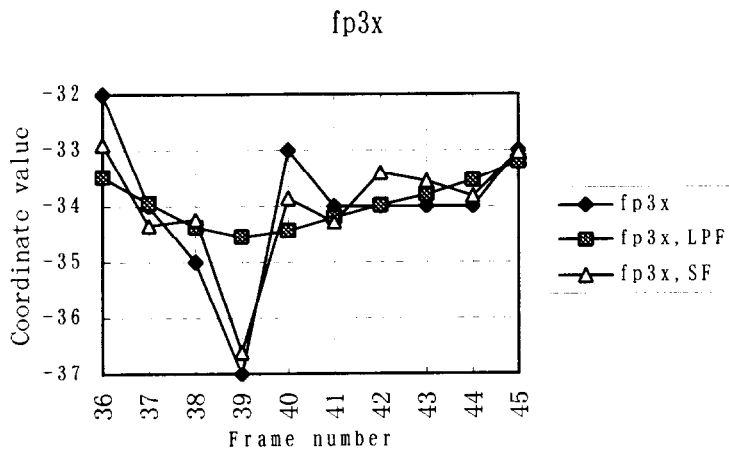


図 4.14(d) 特徴点”3”のX座標値にLPF, SFを6回施した時の比較 (拡大図、縦軸の単位は画素間隔)
 Comparison between X-coordinate values of feature point “3” operated by LPF and SF after 6 times (enlarged)

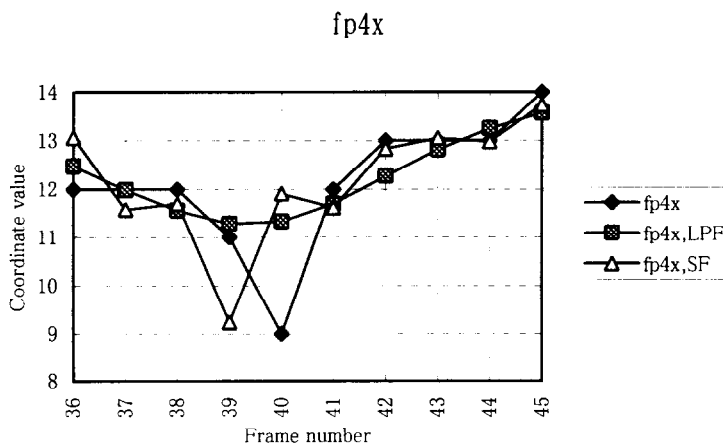


図 4.14(e) 特徴点”4”のX座標値にLPF, SFを6回施した時の比較 (拡大図、縦軸の単位は画素間隔)
 Comparison between X-coordinate values of feature point “4” operated by LPF and SF after 6 times (enlarged)

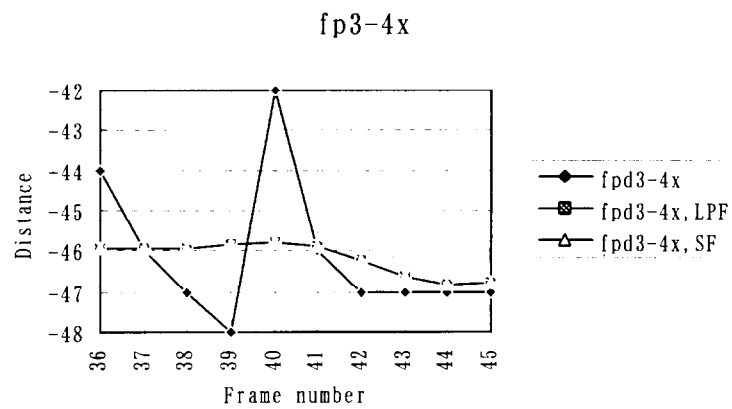


図 4.14(f) 特徴点“3, 4”のX座標値にLPF, SFを6回施した時の3-4間の距離の比較 (拡大図、縦軸の単位は画素間隔)
 Comparison between distances of X-coordinate values of feature point “3 and 4” operated by LPF and SF after 6 times (enlarged)

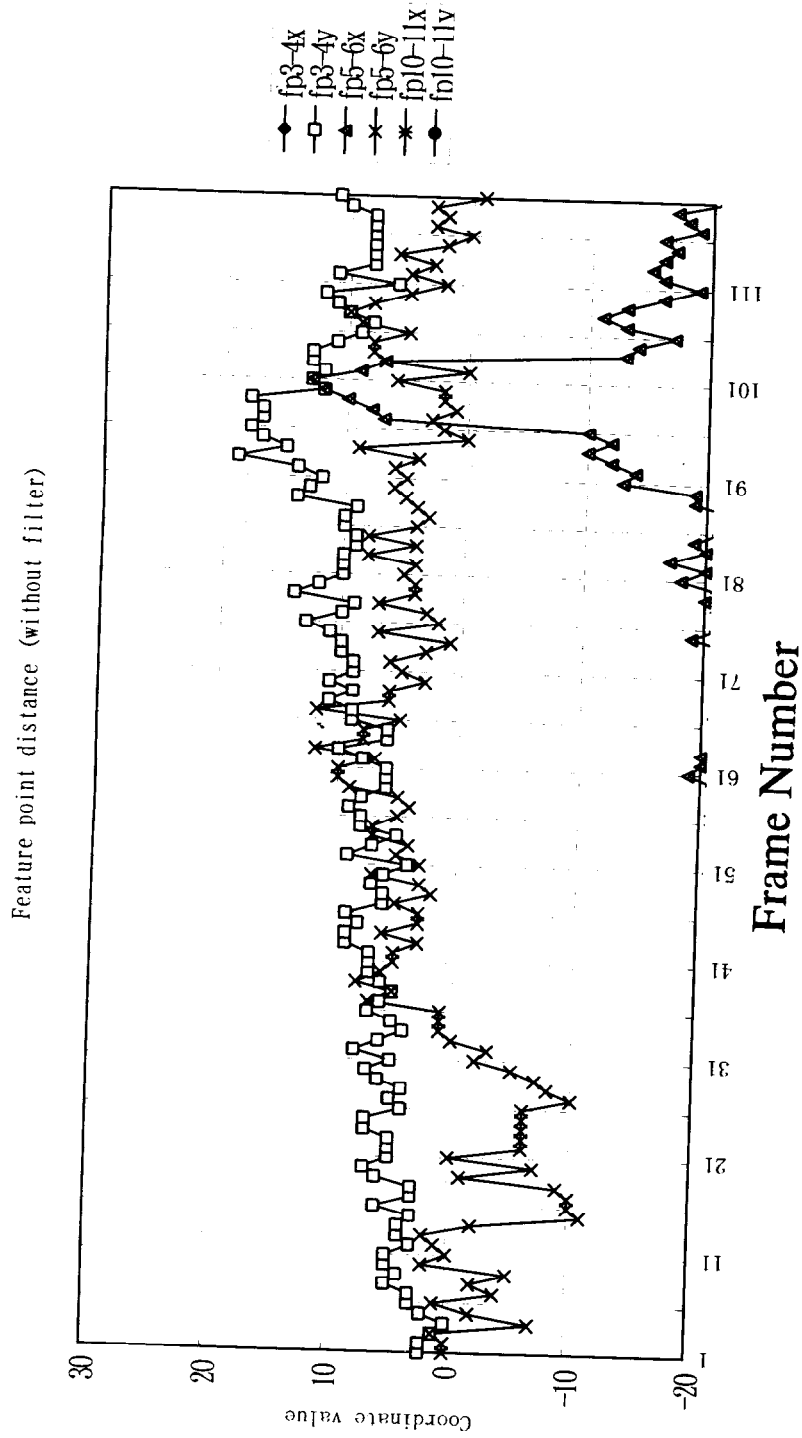


図 4.14(g) 主な特徴点間の距離(フィルタなしの場合、縦軸の単位は画素間隔の差分)
 Distances between several feature points (without filter)

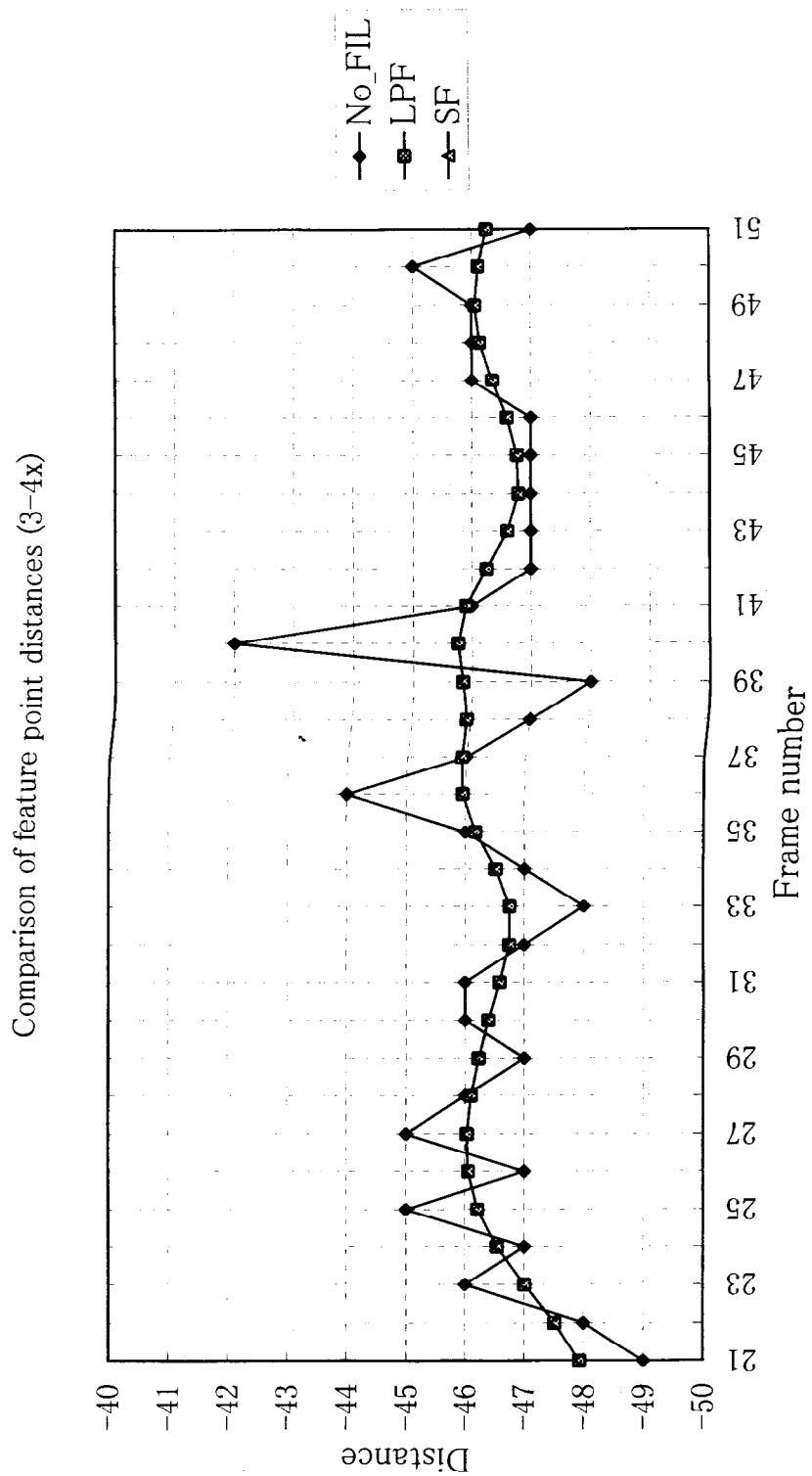


図 4.14(h) 特徴点 3,4 間の距離比較、縦軸の単位は画素間隔の差分
Distances between feature points "3" and "4"

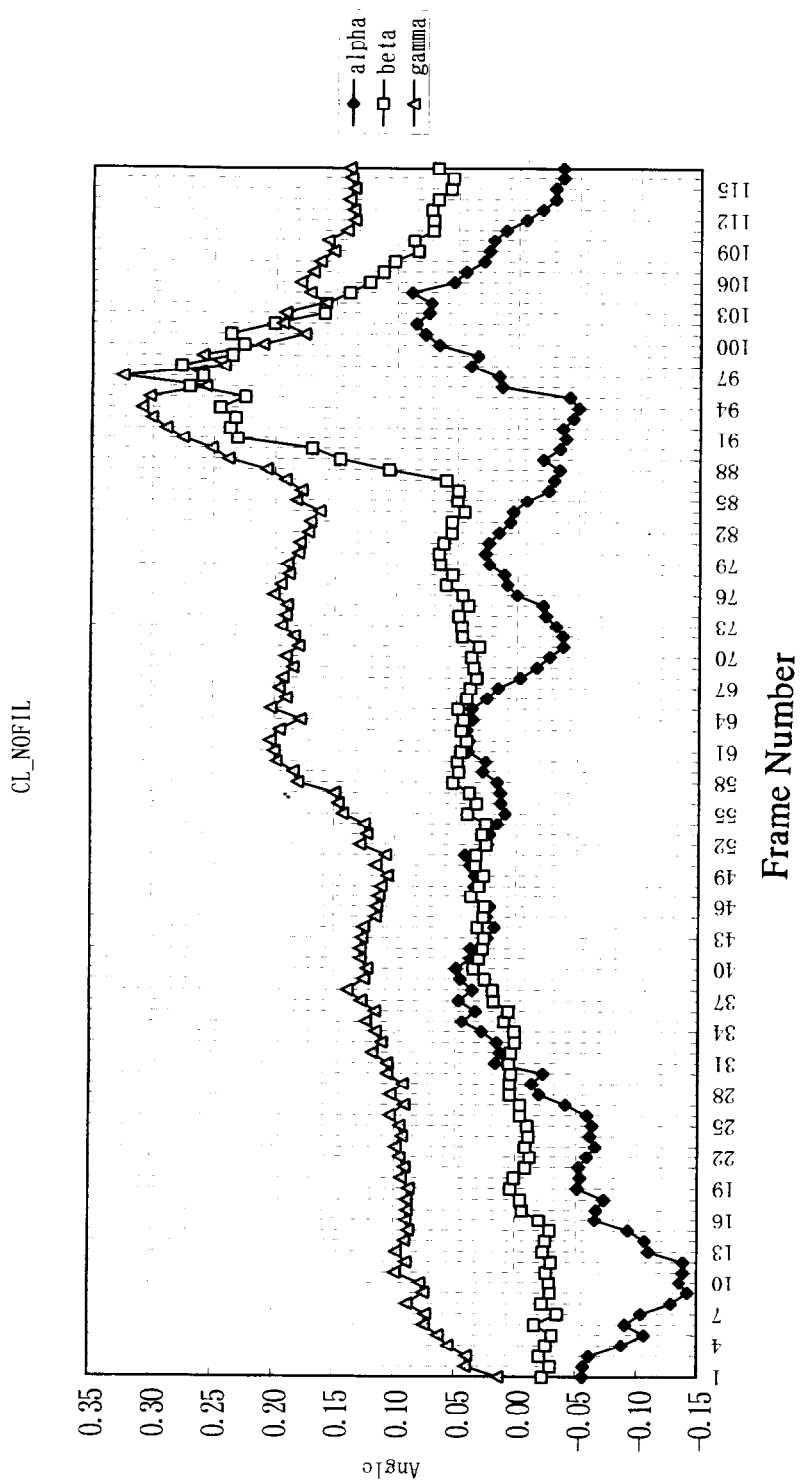


図 4.15(a) 計算されたグローバル動きパラメータ(フィルタなし)
Calculated global motion parameters without filers

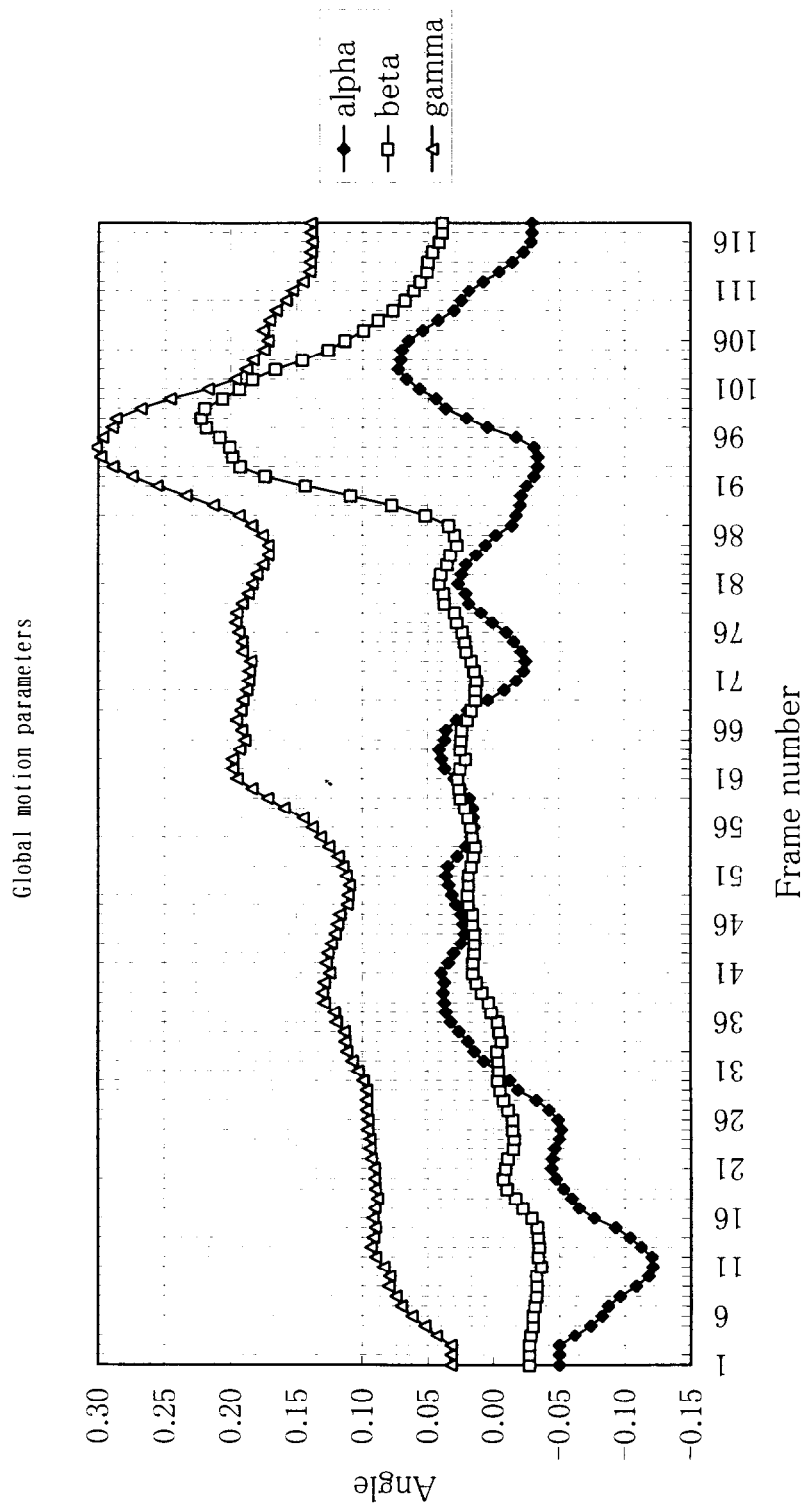


図 4.15(a) 計算されたグローバル動きパラメータ(S F 適用後)
Calculated global motion parameters with Structured filter(SF).

4.8 むすび

インタラクティブ知的符号化の送信側符号化ツールのうち、本章では、頭部グローバル動きパラメータ検出にかかわる部分における問題点とその対策について検討した。インタラクティブ操作の有効性について検討を行い、全自動とマニュアルの間に、人間がどのような入力情報を与えれば、どのような中間的な段階が設定出来るかの検討を行った。GUIを用いて比較の入力し易い情報として顔画像上の特徴点を検討した。その結果、顔画像上に12点の特徴点を定め、全自動とマニュアルの中間的段階として2次元から3次元の動きパラメータを検出する課題を設定した。この課題に対し、顔の3次元ワイヤフレームモデル等の初期条件を加え、奥行きとグローバル動きパラメータを求める2段階からなるアルゴリズムを新たに開発した。動きパラメータの検出は、従来2次元観測データから得られる9個の2次式からなる連立方程式を解かなければならなかったが、特徴点3個の組合わせから求まる3個の連立2次方程式から連立一次方程式を導く解法を求め、より正確な動きパラメータの検出方式が開発できた。線形方程式により、グローバル動きのパラメータ検出を行い、標準画像 Claire に対し良好なワイヤフレームの動きを再現できることを確認した。

インタラクティブ操作は規格化された処理事項を実行すれば、十分な画質が得られるものでなければならない。本章ではフレームごとにほぼ同じ基準で点の位置を指示しているつもりでも、人間の操作には変動が起こるという問題があることを、実験とデータにより明らかにした。グローバル動きパラメータ検出では剛体仮定を行うが、現実には入力特徴点にノイズが付加されたり、ローカル動きの影響があり、剛体仮定に全面的に依存した3次元処理を行うことは難しい。

本章では、まず特徴点の変動をノイズとみなし2次元画面上での顔形状が時間方向に対して局所的には保存されるという観点から、フィルタの構成を見直し、新たに構造的なフィルタを開発した。従来のように時間方向にLPFを施すだけでは、各点ごとに独立したノイズの成分は除去できるが、時間方向に対する動きの周波数成分も減衰してしまう。これに対し、仮想不動点からの特徴点への距離を拘束することにより、特徴点相互の距離の変動を抑圧することができるようになる。この考えに基づき、顔の構造を時間方向に維持するフィルタを開発した。この構造型フィルタは、特徴点の相互の距離の時間的変動を抑圧しワイヤフレームの形状を保持する効果を持つ。ま

第4章 グローバル動き検出とノイズ除去フィルタ

た、本方式では特徴点として、顎以外はローカル動きによる大きい変動の影響を受けない点を用いているため、大きい回転や表情の変化する画像にも適用可能と考えられる。

次に、検出精度に関し、グローバル動きパラメータを求めるために、2つの段階で用いる最小2乗法の適用の仕方に対して、一般的に使われている3つの仮定に基づき推定値の誤差の分散の評価を行い、Dixonのガイドラインに準拠した形に対し、分散は15%しか増加しないことを明らかにした。

これらのノイズ除去と精度検証により、インタラクティブ操作において、位置として3画素、距離変動では数画素までの許容範囲内で特徴点の入力操作を行うという仕様を設定することができる。これは特徴点位置をマウスで指示入力する操作として、操作量、難易度共に操作者に依存しないシステムが構成できることを示している。

第5章 ローカル動き検出、表情、眼鏡画像処理

5.1 まえがき

本章では、顔動画像のインタラクティブ知的符号化において、ローカル動きパラメータ検出と眼鏡画像処理についてインタラクティブ操作を明確化する観点から検討する。インタラクティブ操作の曖昧性について考察を行い、表情などのローカル動き検出と眼鏡処理に関し、その手順の明確化、操作量の評価を行う。目、口などの顔部品のローカルな動きの記述については、顔部品の頂点座標情報を用いる方式から、表情の意味を記述し符号とする方式まで各種の方式がある。表情の意味を記述する方式では、画像から表情を理解して言語化し、再度画像を合成する方式を確立する必要がある。ここではまず上記2通りの方式に関連して、合成処理における手順として、グローバル動きパラメータ(gm と略記する)とローカル動きパラメータ(lm と略記する)との適用順序について検討する。ローカル動きとグローバル動きの記述順序の違いをインタラクティブ操作量の比較により行い、顔部品に応じて異なる記述順序を用いることが有利なことを明確化した。目に関しては瞬きにおける目の開閉度をマウスと GUI を用いて入力して行く。目などのように位置に拘わらず開閉度を判定しやすい顔部品に対しては、アクションユニットによる記述を行う構成をとった。合成においては、目のローカル動きを初期モデルで与え、口のローカル動きは移動後のモデル上で与えることにより非剛体の変形歪みや個人ごとに異なる動きも正確に記述可能となった。

口などのローカルな動きは非剛体の変形歪みも含めて頂点の位置で指示できるように工夫をすることにより、大きい回転などのもとでも正確な口形の記述ができるようにした。またこれを実現するために、グローバル動き検出に基づいてワイヤフレームを変形した後、移動後の状態で、特に口の動きを非剛体歪みと表情を含めたローカル動きとして検出している。フィッティングされたワイヤフレームモデル上にある頂点をマウスによって変形して行く。このようなインタラクティブ操作において、1つの頂点に連動して唇の初期形状を保存したまま、関連ある頂点集合を移動させることによる操作量の削減方式に加え、唇の形状として平坦な直線とうねりのある曲線をモデルに加える拡張を行なっている。これらは周期波形のパターンでモデル化できるものであり、インタラクティブ操作の過程で人間の指示に対応して切替える方式を構成で

きる。ワイヤフレームモデル上の1頂点に対する移動の操作が周辺の頂点に階層的に波及する移動規則を検討し、初期形状とその変形パターンとして0次、0.5次、1次の周期波形を用いた連動規則を開発した。これにより形状合わせの操作量の削減を符号量を増加させることなく実現した。

次に、眼鏡のある顔動画像処理についての検討を行う。知的符号化においては従来、眼鏡のある画像は対象外とされていた。すなわち知的符号化では眼鏡を外した顔画像を初期画像として入力することによって、問題の回避を図っていたともいえる。しかしながら、例えば講演者の顔動画像記録し、符号化することが考えられ、符号化を行なう都合上わざわざ眼鏡を外して初期画像を入力したり、講演中も眼鏡を外したり、講演者から眼鏡のみを借り受けて別途その画像を入力することなどを、いつも行なえるとは限らない。このような想定では眼鏡をかけたままの顔動画像を処理する必要がある。このような応用のためにインタラクティブ知的符号化のツール開発を行い、画像から眼鏡除去の処理を行なう時の操作が曖昧になることが問題であることを述べる。この問題に対応するために、眼鏡消去操作後の残留誤差に定量的評価を定めることを試みる。開発したローカル動き検出ツールを拡張し、眼鏡画像処理を含んだ符号化を行った。眼鏡フレームを肌色で上書きする操作をどの程度まで行なうかの定量的な判断基準として、テレビジョン受信におけるゴースト障害の消え残りレベルとの対応から求まる、DU比を導入した。

本章では、これにより、単眼カメラ入力で得られた実画像に追従したローカル動きを記述することが可能になった。以上のようなアルゴリズムをシステムとして統合し、符号化実験を行い、システムの検証を行った[72]。

5.2 ローカル動き検出と表情の記述

5.2.1 インタラクティブ知的符号化と問題点

図5.1にインタラクティブ知的符号化システムの送信部のブロック構成を示す。顔動画像系列に対し、初期画像1枚と動き等のパラメータにより表現している。初期画像に対しては頭部の3次元ワイヤフレームモデルをフィッティングさせる。図5.2

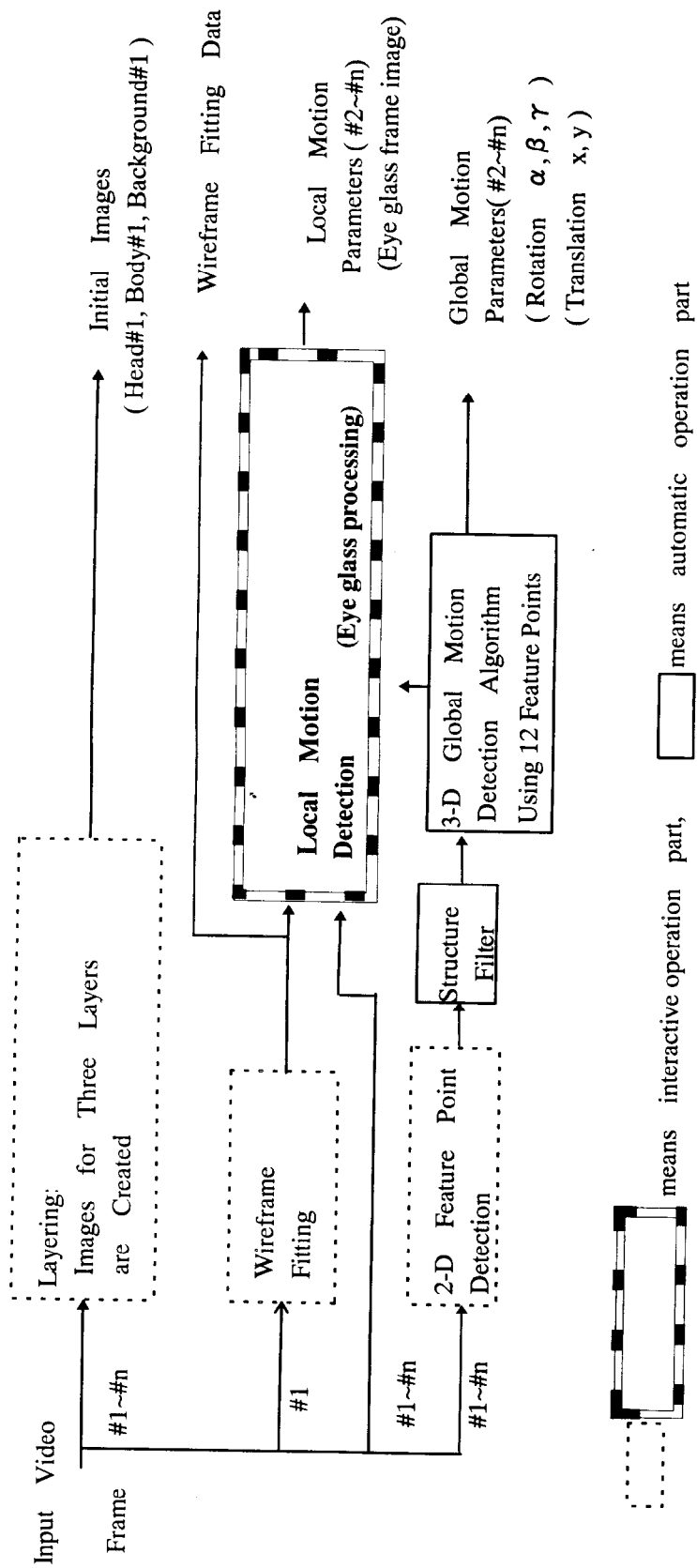


図5.1 ローカル動き検出と眼鏡画像処理を加えたインタラクティブ知的符号化システムの符号化部

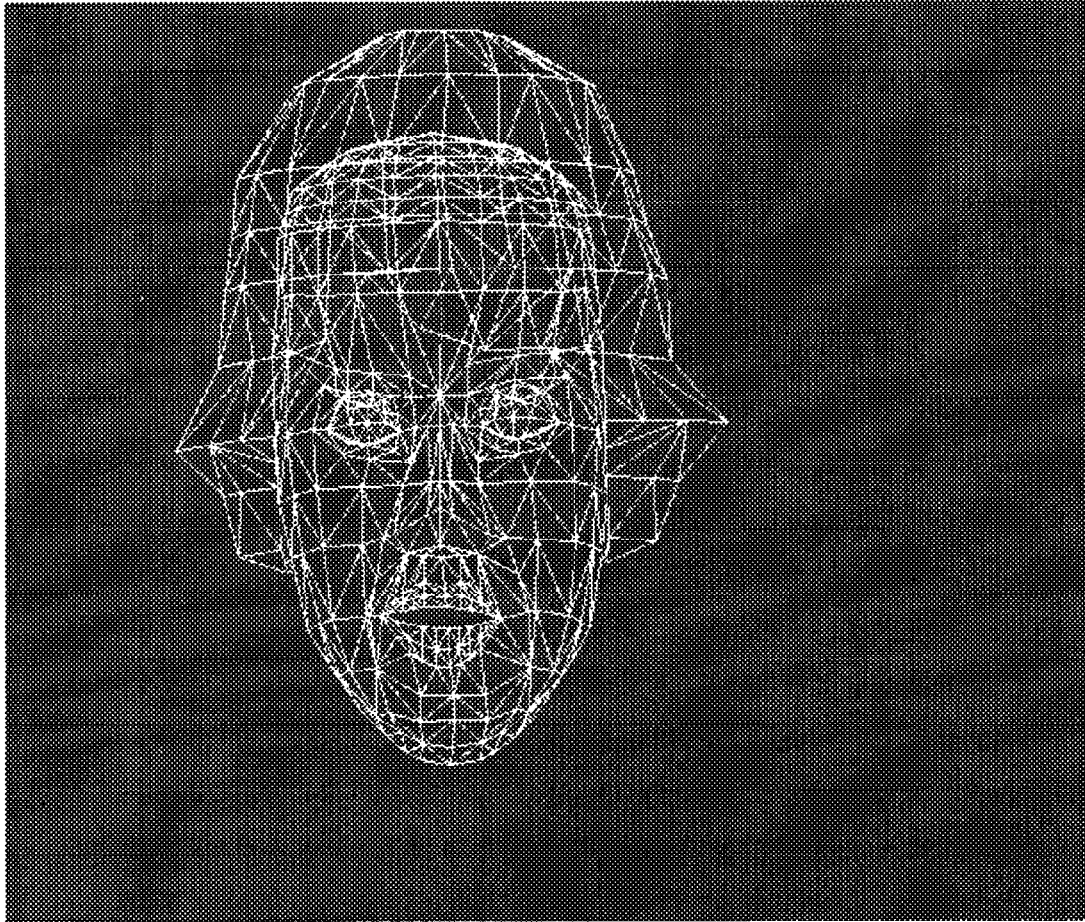


図 5.2 CCITT の Claire(第 3 フレーム)にフィッティングさせた
ワイヤフレームの例

Wireframe model example fitted to CCITT Claire's third frame

に CCITT の Claire にフィッティングさせたワイヤフレームの例を示す。このワイヤフレームモデルに対して回転、平行移動などのグローバルな動きと表情などのローカルな動きによる変形を加え、初期画像のテクスチャマッピングを行い、各フレームの画像合成を行なう。符号化されるパラメータは初期画像、ワイヤフレームモデルのフィッティング情報、グローバル動きパラメータ、ローカル動きパラメータである。ローカル動きパラメータのうち目の形状は瞬きの度合いという一種の意味的なコマンドで表現されている。太い点線で囲まれた部分が本論文で検討している処理である。

インタラクティブ知的符号化の符号化部のうち、初期画像に対する頭部の3次元ワイヤフレームモデルのフィッティング、特徴点入力、ノイズ除去フィルタ、グローバル動きパラメータ検出アルゴリズムについては文献[42][48][73]に述べられている。初期画像に対するワイヤフレームモデルのフィッティングは80の点についてなされるが、顔の輪郭などの約20点は省略できることが多い。グローバル動き検出は顔画像のワイヤフレームモデルを剛体と仮定し、頭部に対する回転と平行移動パラメータを求めている。頭部が剛体に近い場合は初期画像のワイヤフレームモデルを正確なグローバル動きパラメータにより回転・平行移動させれば移動後の顔部品にフィッティングしたワイヤフレームが得られる。しかし実際には発話や表情による変形、また顔自体が非剛体であることによる変形があり、これらを合わせてローカル動きとして再度変形することが必要になる。グローバル動きが線形な動きであるのに対し、ローカル動きパラメータは発話や表情などの非線形な動きも含んでおり、特徴点によってその動きを表現する場合、点数は1つの顔部品に対し10~数10個になってくる。これに対し顔の表情をアクションユニット(以下AUと略す)と呼ばれるパラメータに分解し、表現する手法が研究されて来た[41][74][75]。Ekmanによれば44種のAUにより顔の表情が表現できる[43]。顔画像の合成においてはAUは有効だが、実画像からのAUの検出は正規化や大きい回転などに対する検出に問題が残っている[76]。

従来ワイヤフレームモデルを用いた顔動画の合成においては、初期ワイヤフレームモデルにAUを適用した後、グローバル動きを適用する手法が採られていた。これに対し、筆者らは、グローバル動き検出後に表情変形やその他非線形歪を合わせてローカル動きとして記述することにより、入力実画像に忠実な符号化方式を実現してきた。本章では、その差異を操作量の観点から評価していく。

次に知的符号化では従来より分析合成の都合上眼鏡や髭のある画像の処理は対象外とされていた。これはテレビ電話など実時間応用を前提としていたため、顔画像の特徴点の検出、追跡、動きパラメータ検出などにはまだ困難な部分があるため、その処

理の負荷を増加させる眼鏡画像処理の検討がなされなかったとも考えられる。本章のインタラクティブ知的符号化では非実時間応用であるため、インタラクティブ操作を介して眼鏡のある画像についても動き検出が可能となる。眼鏡画像処理においては眼鏡部分を肌色で上書きする処理と、移動後の眼鏡画像を作り上書きする処理が必要になる。眼鏡画像処理を含む符号化を行ってみることにより、新たな問題点を提起する。肌色での上書き処理は入念に行えば行う程、品質が向上することになるが、眼鏡フレームの境界部などの完全な消去は操作量の増大を招く。実用的なレベルとして、半自動的に例えば数分以内に行える事が望ましい。そこで肌色の上書きにおいてエッジ部分の拡散フィルタの適用や、操作量の上限を設定するため、消え残りに対する定量的評価を与えることなどが必要となる。

5.2.2 ローカル動きパラメータ記述の操作量

図 5.3 は符号化ツールのうち、ローカル動きパラメータ検出に関する部分の動作画面例である。第1フレームに対するワイヤフレームモデルのフィッティングとグローバル動きパラメータの検出は終了しているものとする。第1フレームにフィッティングしたワイヤフレームモデルを検出したグローバル動きパラメータで変換し、現在のフレームの顔画像上にほぼ剛体変形のみワイヤフレームモデルとして上書きした上で、表示する。符号化パラメータとして各フレームで入力した点数、累積点数が表示される。顔部品付近のワイヤフレームモデルのポリゴンの頂点に関しては、発話や表情により上記グローバル動きパラメータを用いて変形されたワイヤフレームモデルと現在のフレームの顔部品との間にずれが生じる。ワイヤフレームモデルの各顔部品付近の頂点位置を移動させることにより、ワイヤフレームの局所的変形を行い、顔部品画像との位置を合わせる。

AUを用いた知的符号化においては、ローカル動き、グローバル動きのパラメータ記述は合成の順序の上で、まずAUで指示されるローカル動き(lm)の変形を行い、次にグローバル動き(gm)を行うという順序(以下lm→gmと略記)で処理が行われて来た。これに対し、実画像入力のlm記述に関しては、記述順序lm→gmよりgm→lmの方[77]が操作量、計算量共に有利であることを表5.1を用いて説明する。

まず、gmは線形な変換であり、ローカル動きがなければ $gm^{-1} * gm = I$ (単位変換)である。変換gmは剛体仮定のもとに行われているが、実際は例えば顎の伸縮など非

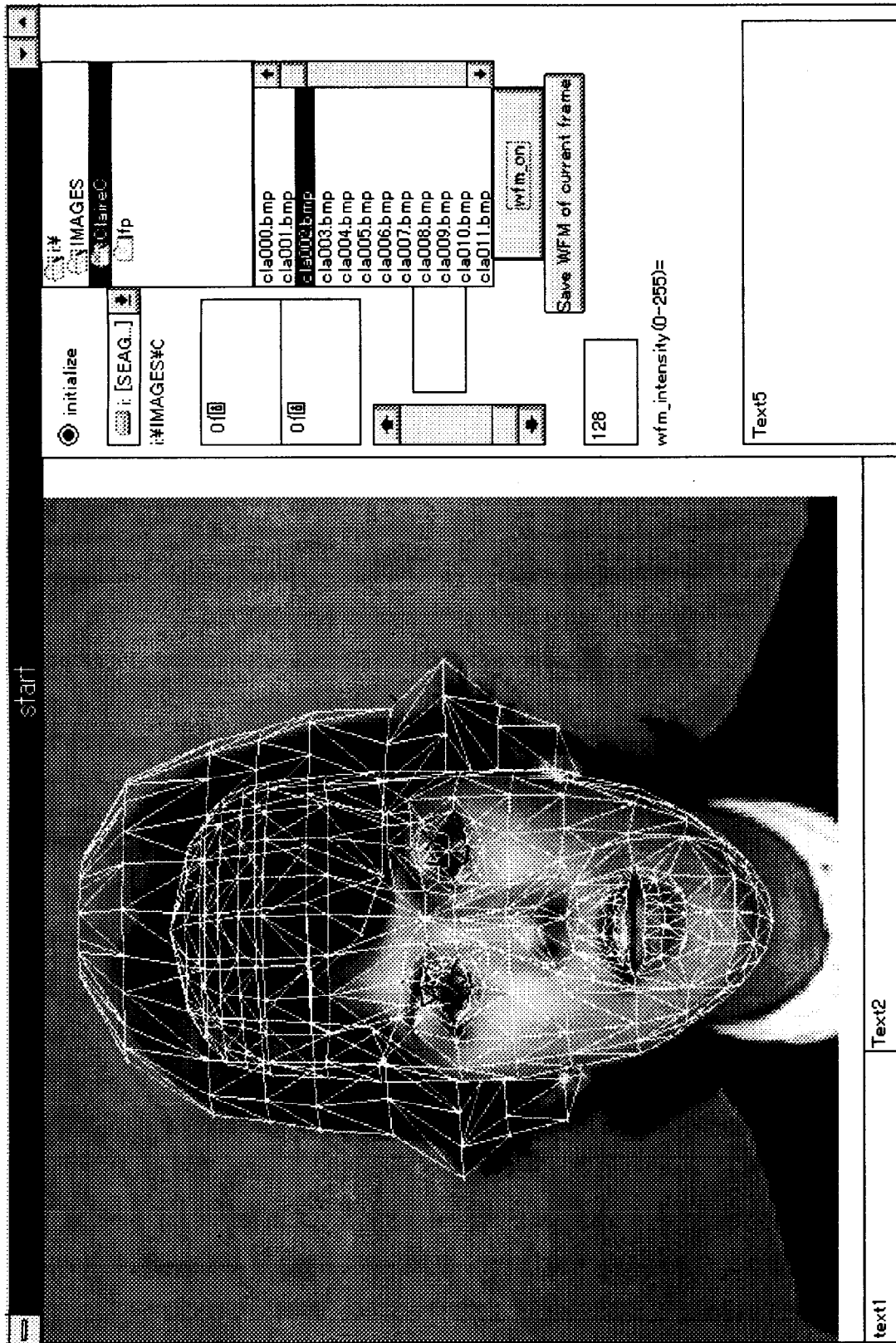


図 5. 3 ローカル動きパラメータ検出を行うインタラクティブ操作ツールの動作画面
 Displayed image in interactive operation tool for local motion parameter detection.

剛体の変形があり、ローカル動き記述でこれらを加味していくと線形の関係がくずれる。lm→gm 方式は初期画像を基準に第 n フレームのワイヤフレームモデルのローカル動きを求めるためには、第 n フレームを基準とした初期画像への対応が必要で、第 n フレームのワイヤフレームモデルフィッティングが新たに必要になる。これはローカル動きを求めるためのものであるから、初期画像のワイヤフレームモデルをグローバル動きパラメータで変換したものより精度の高いフィッティングが必要となる。初期画像に対するワイヤフレームモデルフィッティングの方は、初期画像として処理するシーン全体にわたり最も正面向きで表情の強くないフレームを選択できるため、容易に行える。

表 5.1 ローカル動き記述順序の比較

Comparison of local motion description order.

Description order	lm→ gm	gm→ lm
To get lm	$Tx * gm^{-1}(wfm_n)$	$gm(wfm_0)$
Wireframe Model Fitting	wfm_n (n=0,1,...,N)	wfm_0
gm^{-1} Process	$gm^{-1}(wfm_n)$ (n=1,...,N)	0
synthesis	$Tx * gm * lm(wfm_0)$	$Tx * lm * gm(wfm_0)$

lm:local motion, gm:global motion

Tx: Texture mapping, wfm: wireframe

図 5.4 に示すように、非線形なローカル動きが微少でなくなると、 gm^{-1} 後の顔輪郭や部品等の位置が初期画像のワイヤフレームモデル wfm_0 と違ったものとなる。これは初期画像のワイヤフレームモデルやテクスチャを複数用意しなければならなくなることを意味し、知的符号化の効率化に逆行する。表 5.1 の lm→gm の方式を実現するためにはローカル動きの線形性を確保すべく、ワイヤフレームモデルのフィッティングを行い、その wfm_0 に対するグローバル動きの逆変換 gm^{-1} を実行することなどが必要である。あるいは、別の手段により AU を検出し、その AU で記述するなどの方式も考えられるが、AU 検出は自動処理でもインタラクティブ操作でも困難な処理のひとつになっている。

表 5.1 から各フレームに対するフィッティング作業と gm^{-1} の演算量とが増加した処理量であることが分かる。具体的には CCITT の標準画像 Claire の場合、95-115 フレームの間は大きい回転によりフィッティングの難易度が高くなると予想される。他にも開口時には顎の変形によりフィッティング操作が増加すると予想される。ローカル動きの増加に伴い、図 5.4 に示した gm^{-1} の対応のずれも増大してくる。更に lm は検出時は実数ないしは画素単位の値（整数値）であるが、量子化によって、更に精度が削減される。このとき lm という非線形なデータに対する gm^{-1} での非線形演算としての誤差が加わるため、 lm を直接量子化した場合に比べ、精度が異なってくるという問題もある。

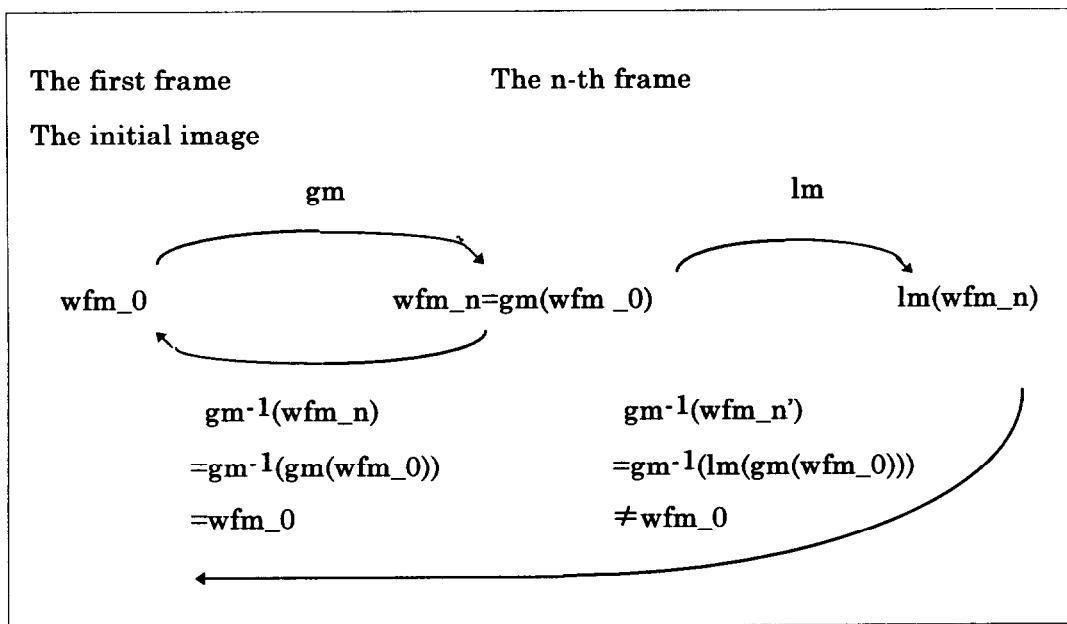


図 5.4 $lm \rightarrow gm$ 記述のために必要な処理

Process needed for $lm \rightarrow gm$ description.

以上から $gm \rightarrow lm$ の記述順序は顔部品でいうと、口の動きやその他表情・変形に適している。一方、目の動きについては、主に短い時間（3~4 フレームの間）の開閉が表現上重要であることから、顔の向きや表情に関わらず開閉度のみを抽出すればよい。この開閉率はそのままロバストな AU となるため、目の開閉に関しては、 $lm \rightarrow gm$ の

記述が上で述べた処理量の増加無しに可能となる。もし目の開閉を gm→lm で記述すると、検出された開閉度から AU を定めて、gm(AU)により開閉後のまぶた周辺のワイヤフレームモデルのポリゴンを変形させるか、各フレームでまぶた周辺のワイヤフレームモデルのポリゴンの多数の頂点を指示入力する必要がある。このいずれの場合も gm→lm 記述に比べ、処理量が増加することになり、AU 検出が容易な顔部品については lm→gm の記述方式の方が有利であることがわかる。以上を表 5.2 に示すように統合し、本論文では目のまばたきに関しては lm→gm 記述とし、口などの AU 検出の複雑性が高いものに対しては gm→lm 記述とするハイブリッド方式として構成した。

表 5.2 ハイブリッドローカル動きパラメータ記述方式

Hybred method of local motion parameter description.

Facial Parts	Eyes	Other than eyes (mouth)
lm discription	AU detection as blinking ratio	lm detection as coordinate values
Description order	lm1 → gm	gm → lm2
Synthesis	Tx*gm*lm1(wfm_0)	Tx*lm2*gm(wfm_0)
Total synthesis	Tx*lm2*gm*lm1(wfm_0)	

5.2.3 周期パターンを用いた連動処理

口形状を構成している頂点の数は、本論文で採用しているワイヤフレームモデルにおいては約 50 点になっている。口付近のワイヤフレームモデルの拡大図を図 5.5(a)に示す。又、図 5.5(b)にワイヤフレームモデルのフィッティングを行う時の口の 6 点を白点で示す。

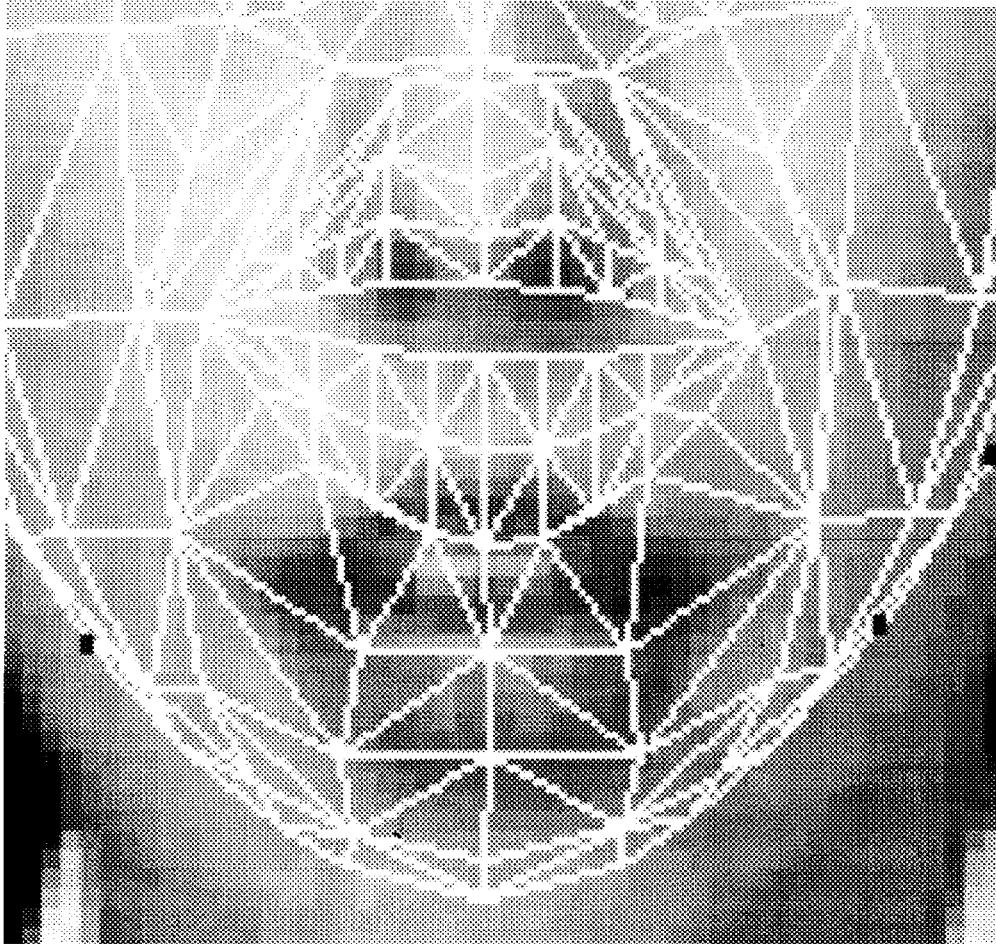


図 5.5(a) 口付近のワイヤフレームモデルの拡大図
Enlarged wireframe model around mouth

口の形状・表情の変化に伴い 50 個の頂点を全てインタラクティブ操作により移動させると操作回数が多く、操作者への負担が多い。発話・表情に伴う動きの多くは、ポリゴンの頂点が連動して移動している部分が多い。そこで、口形状を形成するポリゴンの頂点を階層的に定義し、上位階層の頂点が移動した場合はそれ以下の階層の頂点も連動して動くようにし、一様な動きに対しては指示操作の手間を省略できるようにする。ここでは連動に際して、復号器側の合成処理における演算量が増加しないように、内挿比率を乗算する程度の処理で階層処理方式を構築した。

口の付近のワイヤフレームモデルの連動規則を作成するために、口形状の動きのパターンをモデル化することを考える。まず、口形状のモデル化の例として、唇の輪郭上の 8 点の座標を制御点とし、残りの点をスプライン曲線で近似し、更に丸い形状に対する特殊処理を行う手法がある[78]。これはモデル化の段階で 1 回行うものであるため、指示点数が多く、本方式には不適である。この場合の指示点数の削減は操作量の削減とともに、符号量の削減にもなる。そこで本論文では、口の両端 2 点を結ぶ線分に対し、周波数成分に対応したパターンによる表現を行なう 0~1 次の周期波形をモデル化し、操作量の削減を試みる。0 次は平坦な線分、0.5 次は弓なりの半周期の円弧、1 次は 1 周期のうねりに対応するものを考える。

階層の順序に従って上位から階層ごとに移動量を求め、その階層の全頂点の処理が終了した後、次の階層の処理を行う。以下繰り返し処理を行い、下位階層がなくなるまで処理を行う。上位階層の頂点を親とし、下位階層の所定の頂点が子として移動されるが、この親子関係を頂点ごとに後述の手順で定義しておく。親の移動に伴い、まずその子がサーチされ、同時にその子に対する 2 親の移動後の位置がメモリから呼び出される。次に図 5.6 に示すように、2 親の移動前の点の位置からの移動後の 2 個の移動ベクトルを内分比率に従って合成することにより子の移動量ベクトルが計算される。この移動量ベクトルが子の移動前の位置に加えられ子が移動される。この処理は巡回的になされるが親子とも移動前の位置はそのフレームの処理を開始する前の位置を用いている。この処理により上位階層の移動に起因する下位階層の形状は、そのフレームの処理を開始する前の始めの状態が保たれる。従って、口形状の場合、上下方向の開閉や平行移動の場合は、上位階層の数点を指示するだけで他の頂点は連動するため、指示する必要がなくなる。更に口部分に独特な形状・表情や歪みなどがある場合は、下位階層にまで及んで指示を行うこともできる。

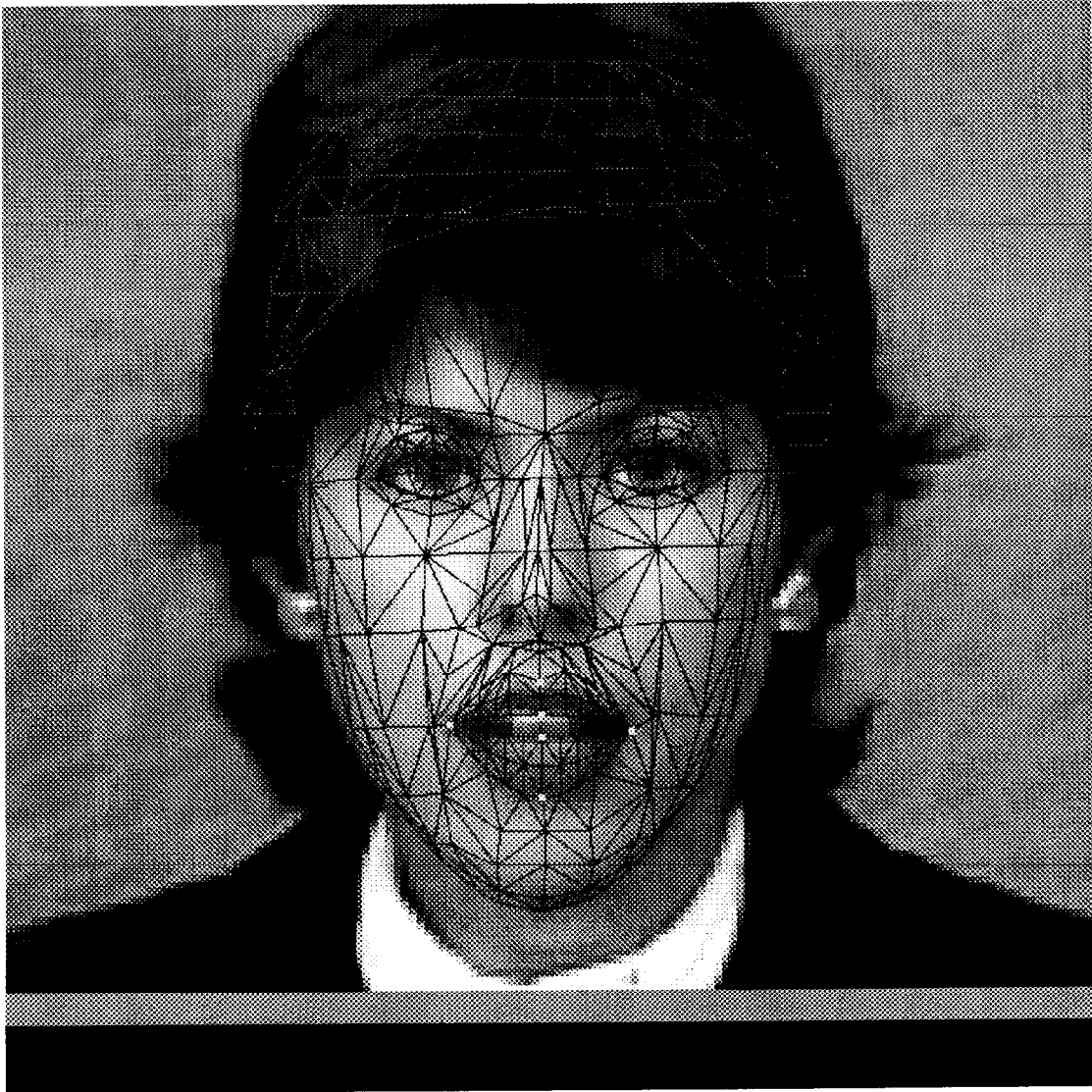


図 5.5(b) CCITT の Claire に対するワイヤフレームモデル
(口は6点で初期画像にフィッティングされる)

上のような定義に従って、連動的な動きをする場合と単純な線形内挿との比較を図5.7に示す。(a)は片方の親 P_1 の P_1' への移動に伴い子 C が C' へ移動した場合で(b)は更にもう一方の親 P_2 が P_2' へ移動し、全体として子 C'' へ移動する様子を示している。(c)、(d)は同様の親の移動に伴い、子が線形内挿される様子を示す。式では

$$C_{\text{new}} = C_{\text{old}} + \alpha \overrightarrow{P_1 P_1'} + \beta \overrightarrow{P_2 P_2'}, \quad \alpha + \beta = 1, \alpha, \beta \geq 0 \quad (5.1)$$

となる。ここで、 α 、 β は2親の影響力を与える内挿比率である。

5.2.4 連動規則の登録方式

前節の連動処理を行なうためには親子関係の規則を予め登録しておかなければならない。以下その手順を示す。まず、顔画像上にワイヤフレームモデルを配置し、次に連動させる頂点を選択する。選択された頂点に対し、その頂点に影響を与えるのに適している親となる頂点を2点選択し、登録を行う。口形状の場合は口の両端を最上位の階層とし、その中点を次の階層の点とし、更にその中間の点を下位の階層の子として登録していく。また口形状が開口により広がるため、口から顔輪郭の間のワイヤフレームモデルの頂点に対しても、同じく親子関係の定義を行い、口周辺の画像のバランスを保つようにしている。また親から子を内挿する際の比率 α 、 β は x 、 y 座標ごとに2親と子の距離の比を用いた。

5.2.5 連動と操作量

以上の規則を定めておくことにより、通常の子の動きに対し、操作の指示回数は4~6位で、ほぼ原画に重ねる事ができた。Claireの105フレームのローカル動きパラメータの形成過程を示したのが図5.8である。この階層的な移動規則は口の輪郭内の頂点のみではなく、顔輪郭にまで波及し、周辺のテクスチャに皺ができないようになっている。口付近で指示回数の多かった頂点は、口の両端と上唇下唇の中点の上下端であった。図5.9は口が閉じた場合で、初期形状が弓型になっているので、直線状にするため指示点数は増加する。この場合、上下の唇を互いに重ねるように配置することにより、指示点数を節約して、閉口を表すこともできる。

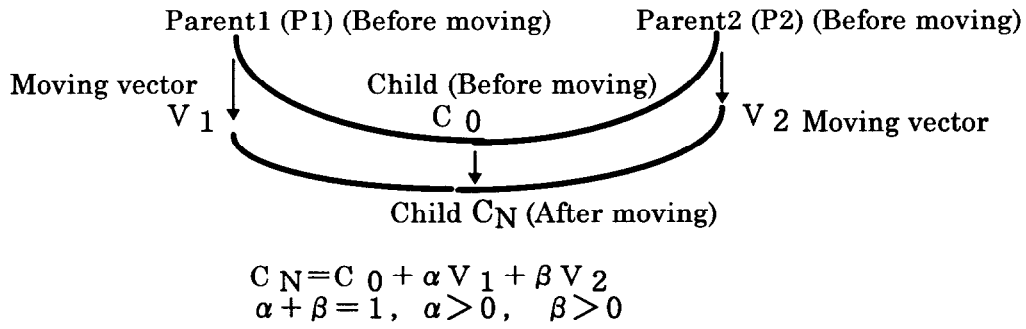


図 5.6 親子の点の移動関係
Relation of parents and child moving.

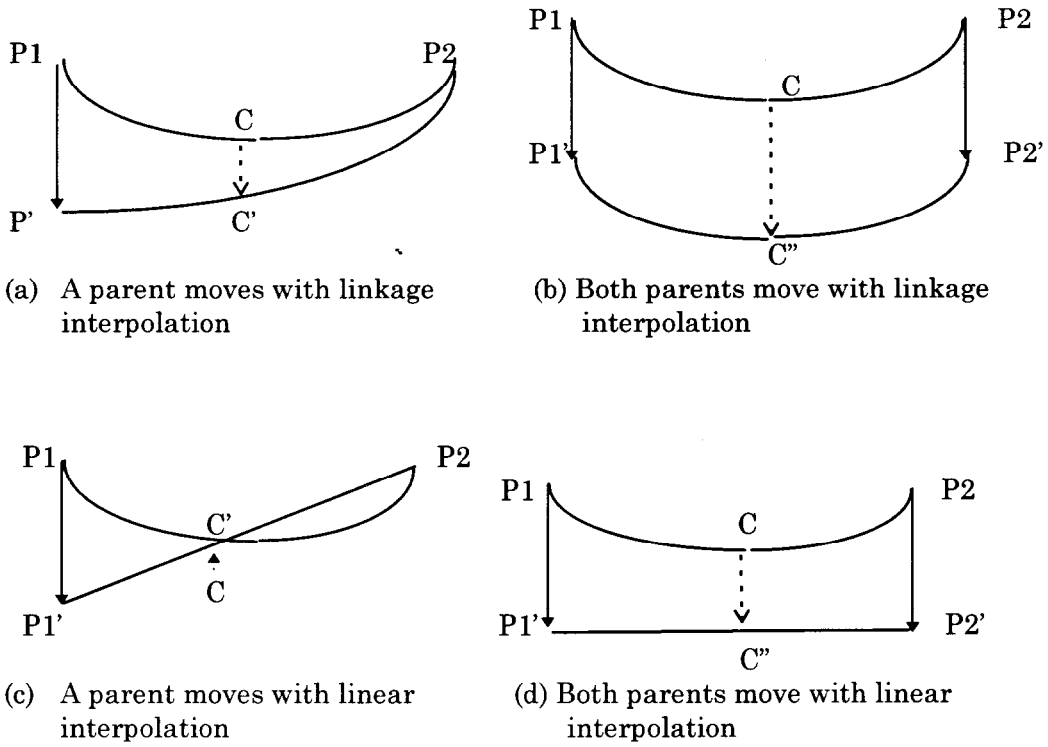


図 5.7 連動的内挿と線形内挿の比較
Comparison between linkage interpolation and linear interpolation.

図 5.10 は CCITT 標準画像の別の系列 Miss America の第 101 フレームの画像である。唇の形が発話によりうねりを持った形になっており、特に唇の内側の輪郭の形状の入力を行う必要がある。

次に具体的に連動の動作について説明する。0次である閉口の形状の時、唇の両端点（図 5.11(a)の q_1, q_2 ）の修正を必要に応じて行う。グローバル動き変換後、両端点にずれがない時は修正はしなくてよい。次に閉口においては唇の midpoint q_3 （又は q_4 ）の位置を指定するが、これが q_1, q_2 を結ぶ線分の midpoint m_3 との中間より m_3 側に近いとき、 q_3 を q_3' に移動させ、 q_1, q_2 を親とする連動関係を移動後の親の線形内挿

$$C_{new} = C_{old} + \alpha q_1' + \beta q_2' \quad (5.2)$$

で行う。これにより q_3' の 1 回の入力で、上唇の下側の輪郭線が閉口に合わせて平坦に並ぶことになる。下唇も同様である。受信側での合成処理も同じ判断基準で行えるため、これによるオーバーヘッド情報は不要である。但し、閉口に類似して更に微妙な変形を表現しようとして、 q_1 又は q_2 と q_3' の midpoint が更に入力された場合は、その点は更新され、その点を親とする子の点は以下のうねりのモードの場合を除いて式(5.2)で内挿されるものとする。

次に図 5.11(b)に、1次の波形のうねりのモードを示す。点 q_5 の指示において、その位置が q_2, q_4 で決定される位置 q_5' と(5.1)と(5.2)で決定される位置の二つの内挿点の midpoint との距離が q_5' に近い場合は通常モードとして、そうでない場合はうねりモードとして処理される。 q_4', q_5', q_2' で曲線を形成し、途中の子の点の内挿がなされる。親 q_1', q_4' の内挿は(5.3)式により q_6' の位置に設定し、 q_1', q_6', q_4' で曲線が形成される。

$$p_6' = (q_1' + q_4')/2 - (q_5' - (q_4 + q_2)/2) \quad (5.3)$$

以上のうねりモードの処理における上記判定基準は受信側の合成過程でも同一の処理ができるため、オーバーヘッド情報は不要となる。

512x480 画素の顔画面上で、原画に重ねたワイヤフレームモデルが目視でほぼ一致した時点、画素にして 1 画素程度のずれに収まった時点を操作終了の基準としている。但し、再生画像のサイズが小さい場合は縮小比率の逆数の画素分のずれまで許容できるものとする。図 5.12 に 0 次、1 次のパターンを導入した場合のワイヤフレームモデルの形状を示す。図 5.12(a)は図 5.9 と同じ Claire の 46 フレームを(1)式の手法で処

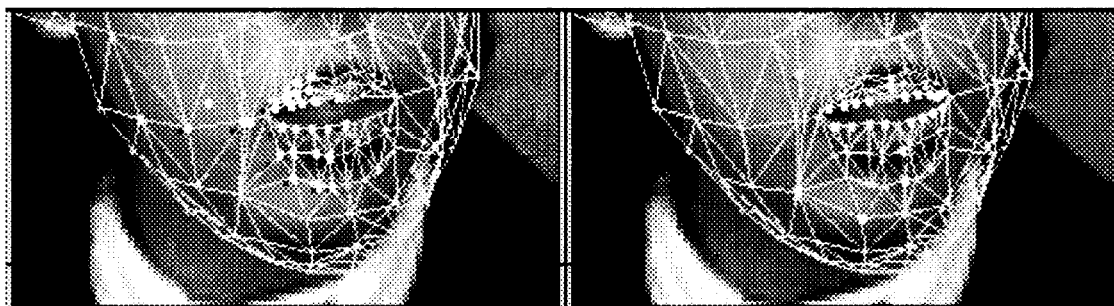


図 5.8 (a) 反復 1 回目

図 5.8 (b) 反復 2 回目

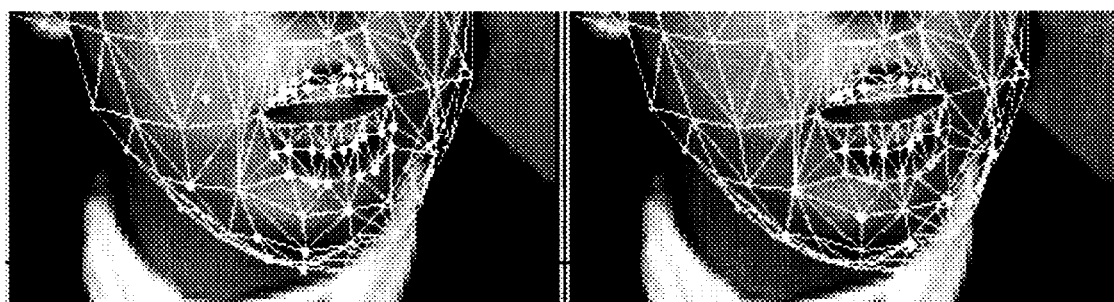


図 5.8 (c) 反復 3 回目

図 5.8 (d) 反復 4 回目

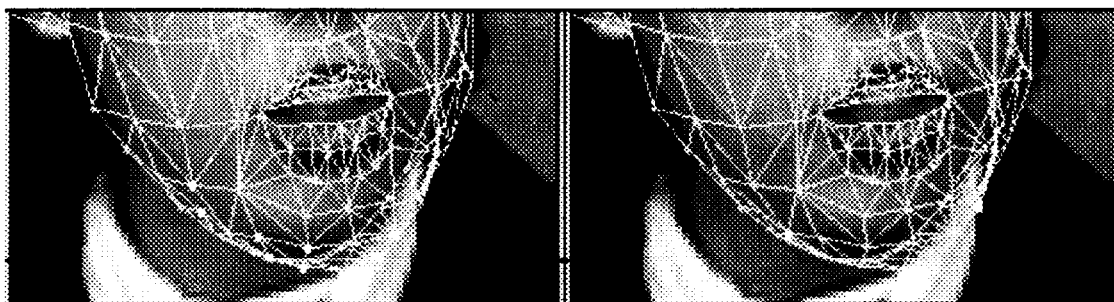


図 5.8 (e) 反復 5 回目

図 5.8 (f) 反復 6 回目

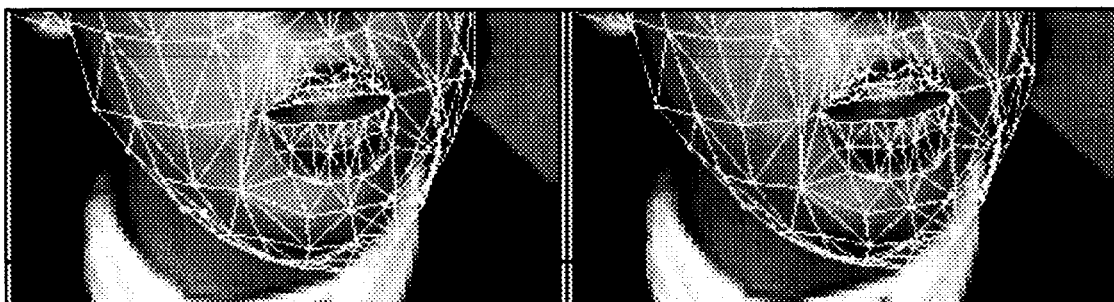


図 5.8 (g) 反復 7 回目

図 5.8 (h) 反復 8 回目

図 5.8 ローカル動きパラメータの座標入力と口周辺の頂点の連動
Input coordinate positions of local motion parameters and
sequential moving of veteces around mouth.



図 5.9 口を閉じた画像 (Claire46 フレーム)

Closed Mouth (CCITT Claire, 46-th Frame).



図 5.10 口形状にうねりのある画像 (Miss America 101 フレーム)

Mouth Image with Waving Shape (CCITT Miss America, 101-th Frame).

理を行った場合で、ローカル動きパラメータを 11 個使用している。(b)は(2)式の手法の場合で、ローカル動きパラメータは 4 個である。(c)は図 5.10 と同じ Miss America の 101 フレームを(5.1)式の手法で処理した場合で、うねりを再現するためローカル動きパラメータを 18 個使用している。(d)は(5.3)式の手法で、ローカル動きパラメータは 6 個で表現されている。

図 5.13 に Claire についてのローカル動きパラメータのフレームごとの個数を示す。表 5.3 に画像 Claire、Miss America、Naomi につき 5 種の符号化時のローカル動きパラメータ符号数を示す。口のワイヤフレームの頂点数 50 に比べ操作量、符号量が共に大きく削減されたのがわかる。操作量としては、平均 2.6~8.2 個/フレームに削減されている。出現頻度の高いフレームのうち 44~48 フレーム付近は唇の中心部が閉口により水平になっているためと考えられ、0 次周期波形のパターン化によりこの部分は 1/3 にできた。

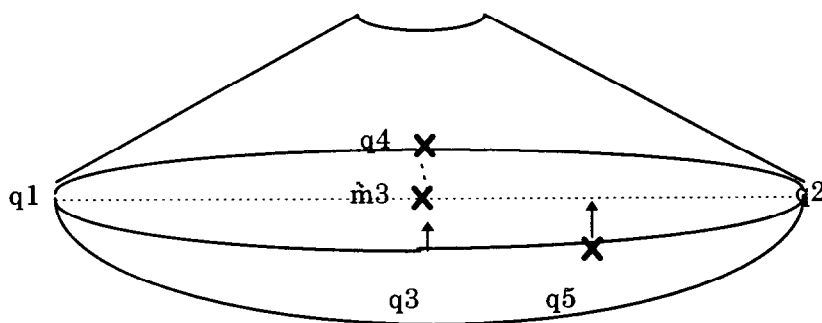


図 5.11(a) 平坦モード時の頂点位置

Vertecies Arrangement for Horizontal Mode.

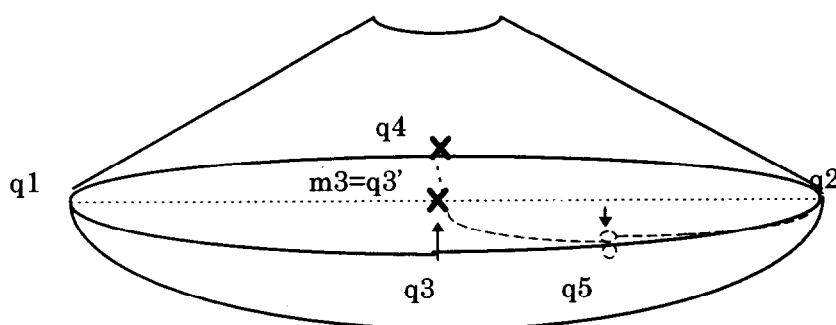


図 5.11(b) うねりモード時の頂点位置

Vertecies Arrangement for Wave Mode

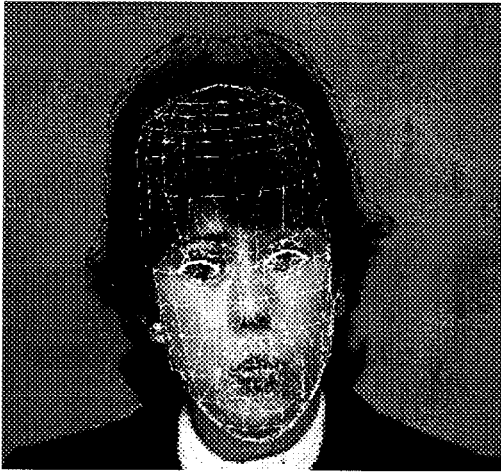


図 5.12(a) 線形内挿の場合
(CCITT Claire 第 46 フレーム)

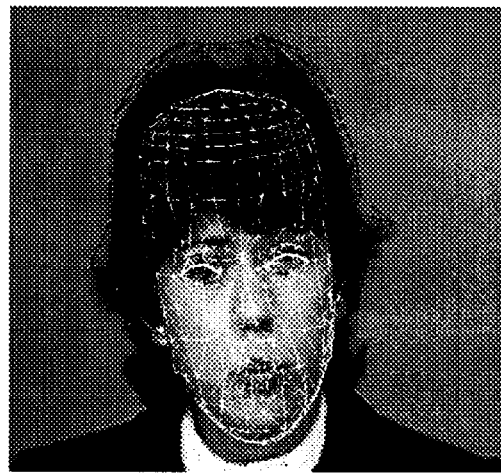


図 5.12(b) 0次パターンを用いた場合



図 5.12(c) 線形内挿の場合
(CCITT Miss America 第 101 フレーム)



図 5.12(d) 0次パターンを用いた場合

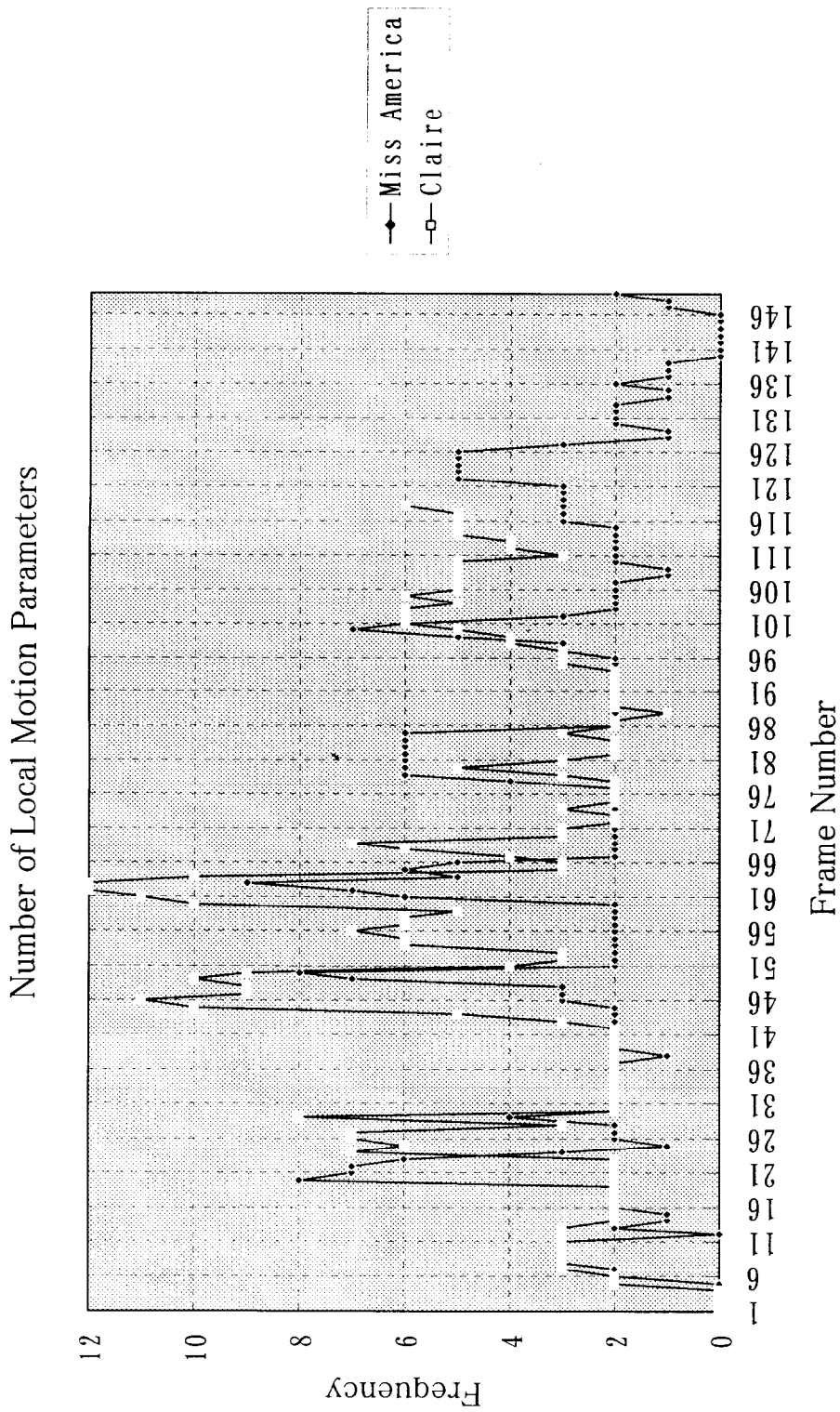


図5.13 ローカル動きパラメータの符号化数 (CCITT Claire)

Number of Local Motion Parameters (CCITT Claire)

表 5.3 ローカル動きパラメータの平均個数

(S=再生画像サイズ 128x120, M=256x240)

Average Number of Local Motion Parameters

(S=Reconstruction image size is 128x120, M=256x240)

Sequence	Claire1 (S)	Claire2 (M)	Miss America (S)	Naomi1 (S)	Naomi2 (S)
Number of Frames	1 1 8	1 1 8	1 4 9	6 0	6 9
Average Number of Local Motion Parameters	3. 9 8 3	8. 2 3 7	2. 6 2 4	4. 1 8 3	3. 3 4 8

5.3 操作量を考慮した眼鏡画像処理

5.3.1 眼鏡処理を行なわない場合の問題点

眼鏡画像処理は従来の実時間処理を前提とした知的符号化においては行われることはなかった。システム全体の中で特徴点の検出、追跡や3次元動きパラメータの検出などの困難であり、眼鏡のない人物顔画像であっても厳密には無条件での動作は難しかった。汎用性確保にあたり、ほぼ正面向きであること、大きな回転は無い、背景に紛らわしい物は無いなど、いくつかの前提条件を設定する必要があった[40][79]。従ってたとえ初期画像に対して、眼鏡部分の検出処理等を施すことができたとしても、それ以後実時間で入力される眼鏡付き画像から正しい動き検出を行う事は、処理の負荷の増加となるため、眼鏡付きの画像処理は行なわれていなかったと考えられる。これに対し、筆者らの提案してきたインタラクティブ知的符号化は、いくつかのインタラクティブ操作事項とその回数、精度などを規定しておき、カメラと被写体のおよその配置という拘束条件を加えることにより、汎用的に動きパラメータを検出するアルゴリズムを開発して来たため、従来行われていなかった眼鏡処理などの拡張は、動き検出とは独立に可能となった。ここでは実際に眼鏡画像処理を含む知的符号化を実行し、そこに生じる新たな問題を調べる。ポリゴンの頂点を移動させることにより、非剛体としての変形もするワイヤフレームモデル上に剛体として変形しない眼鏡フレームモデルを付加した形での処理を行う。

まず、眼鏡付き顔画像を眼鏡処理なしに直接知的符号化した場合について述べる。図 5.14(a)は眼鏡付きの初期画像、図 5.14(b)は AU による表情変形を与えた場合の結果で、目の周辺に表情による変形があるため、剛体であるはずの眼鏡が変形している様子がわかる。

5.3.2 眼鏡画像処理

眼鏡なしの顔画像を入力するためには入力画像系列を撮影する際に被写体となる人物に眼鏡を外してもらい、初期画像の入力を行い、また眼鏡を借りてモデル化に必要な入力を行うという作業が必要である。更に眼鏡のみ別系統で入力する場合、照明条

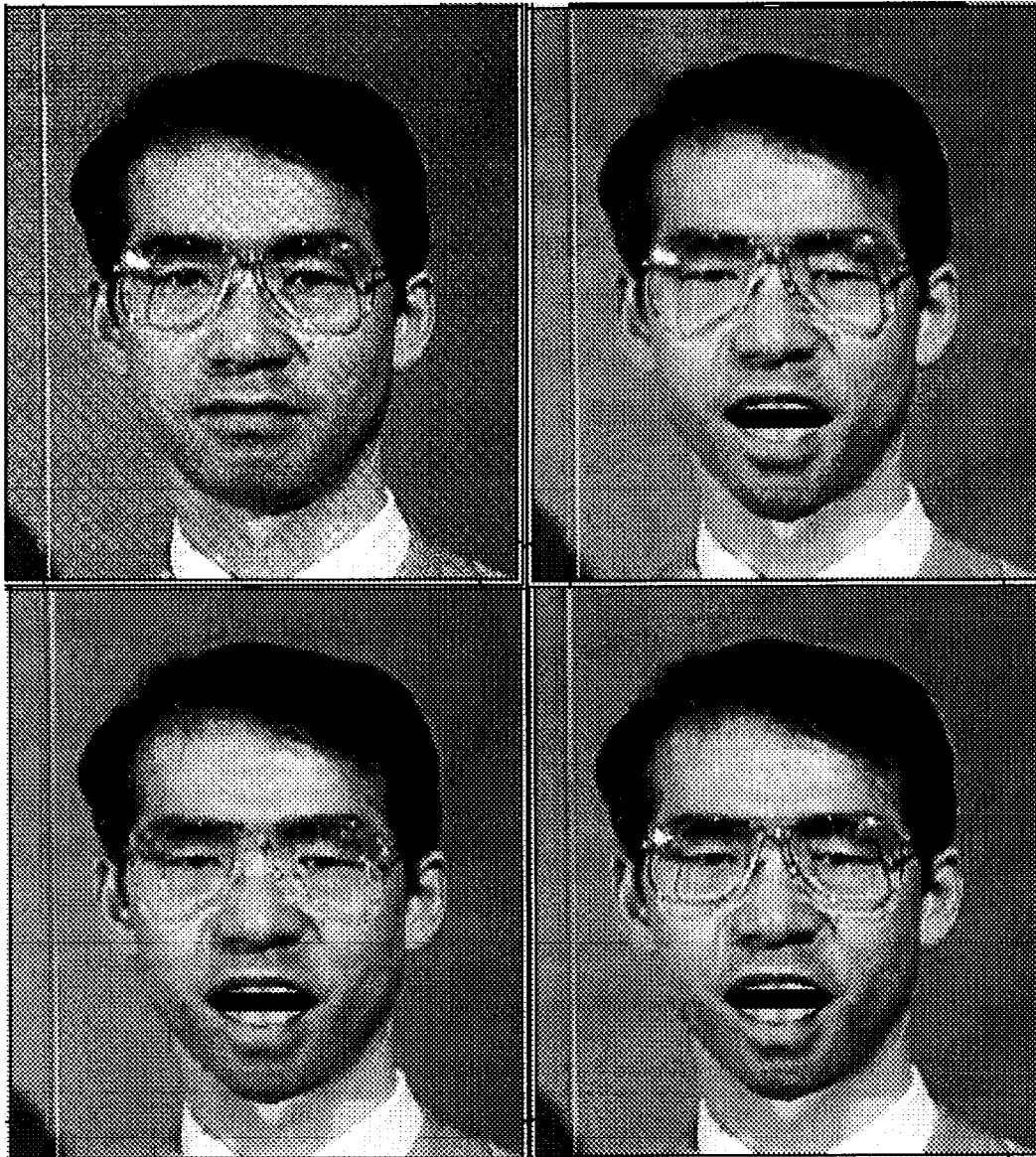


図 5.14 眼鏡画像処理,

(a)左上：原画,

(b)右上：眼鏡画像処理なしの合成画像

(c)左下：眼鏡除去処理結果,

(d)右下：眼鏡画像処理付きの合成画像

Eye glass processing,

(a)Upper left: Original image,

(b) Upper right : Synthesized image without eye glass processing,

(c) Lower left: Eye glass removal result,

(d) :Lower right: Synthesized image with eye glass processing.

件を人物撮影時とほぼ同じ状態にしておくことが望ましい。合成後の上書きにおいて眼鏡フレームと顔画像の境界が不自然に強調されたものになるという、照明条件が異なる眼鏡画像を別系統で入力する処理例として、眼鏡フレームの選択のために仮想的に試着するメガネオーダメイドシステムがある[80]。実際にメガネを装着できない状況で、そのイメージを視察するのに役立つと思われるが、やはり若干眼鏡エッジが強調されている。図 5.15 に眼鏡処理方式の構成を示す。眼鏡処理では、眼鏡フレームやレンズの反射によるハイライト部分を対応する肌色で上書きして眼鏡を除くとともに、眼鏡フレーム部分を切り出し眼鏡フレーム画像として保存しておく。眼鏡除去画像を初期画として動きパラメータによりワイヤフレームモデルに変形を与え、初期画像をテクスチャマッピングした後、眼鏡フレーム画像を上書きする。このように眼鏡付き画像を入力として処理する方式に対し、眼鏡の無い顔画像を入力し、かつ眼鏡は別系統で入力し、テクスチャマッピングによる顔画像合成後、眼鏡部分を上書きすることも可能である。この結果を図 5.14(c)(d)に示す[71][81][82]。

表 5.4 に眼鏡画像処理における上記2種の方式の比較をまとめてある。今後眼鏡をかけた人物の講演などをビデオテープに収録したものを符号化するような応用も多くなると考えられる。符号化対象である人物が眼鏡を外した顔画像を初期画像としてわざわざ入力したり、眼鏡のみを取り外して撮影したりすることに協力的な場合はB方式でも符号化が可能だが、実際はそうでない場合も想定されるため、A方式による符号化が必要となる。

表 5.4 眼鏡画像処理の方式比較

Comparison of Eye-Glasses Processing Methods.

	Method A Proposed method	Method B Separated input of eye-glass frame
Input	Facial image with Eye-glasses	Facial image and eye-glasses separatedly transferrered
Eye-glasses modeling	From multi-frame images	From multi-view images
Motion Detection	From synthesized vertices of wireframe	
Texture adjustment	LPF for edge	Color adjustment for eye-glass frame Naturalization of edge
Synthesized image	Natural texture	Torelable for large motion

第5章 ローカル動き検出、表情、眼鏡画像処理

眼鏡処理における上書きする領域の（端）点、領域の接続、閉曲線内の塗りつぶしにより、眼鏡フレーム部、ハイライト部、かげ部など眼鏡の存在に起因する画像の変化部を肌色または眉などの色で置換えを行う。この置換えはその部分の肌色に近いと思われる部分を数画素サイズのブロックとしてコピーし、上書きする。これは肌色部分は同一色が連続しているのではなく、肌の細かい模様がテクスチャとして存在するため、単一色や類似の1色を選択し置換すると、肌目が失われて光沢が出たりするためである。更に置換え画像は類似した肌色のブロックで上書きした場合でも実際の肌色そのものは未知であるため、眼鏡フレームのエッジ状の痕跡を完全に消すには相当な

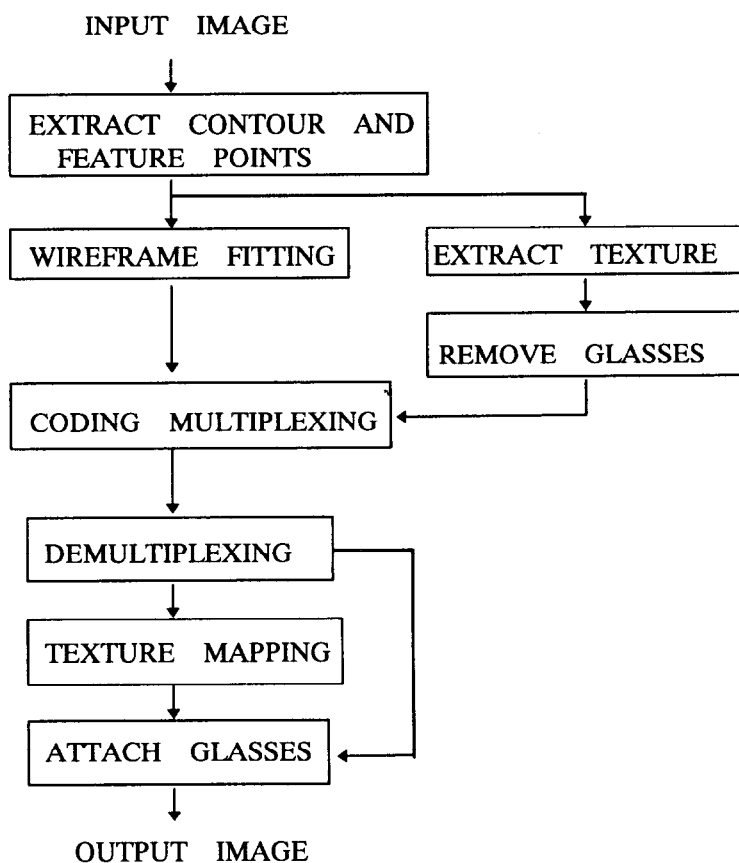


図 5.15 眼鏡画像処理の流れ図
Flowchart of eye glass processing.

第5章 ローカル動き検出、表情、眼鏡画像処理

回数の試行作業を要することがある。ここではまず上書き後のエッジ部分に拡散処理を行い、残留するエッジ成分の軽減を図った。また、後述するように、眼鏡を上書きした時点での消え残りとの眼鏡の輝度振幅の比較から、操作を終了する基準の設定を行った。

眼鏡の動きに関しては

- ①両目の中間点からの z 方向に関する眼鏡の距離を d とし、眼鏡フレームを3次元モデルとして扱う。
- ②グローバル動きパラメータから眼鏡フレーム部分の3次元スケーリング処理を行う。
- ③グローバル動きパラメータから眼鏡フレーム部分の2次元スケーリング処理を行う。
- ④実画像の眼鏡両端部の位置座標から、2次元スケーリング処理を行う。

などの方式が考えられる。①は表 5.4 のB方式に対応し、眼鏡フレームの3次元モデルが必要となる。②~④はスケーリングのみを行う方式であり、その計算手法が異なるものである。ここでは各画像フレームにおけるワイヤフレームモデルの頂点座標が

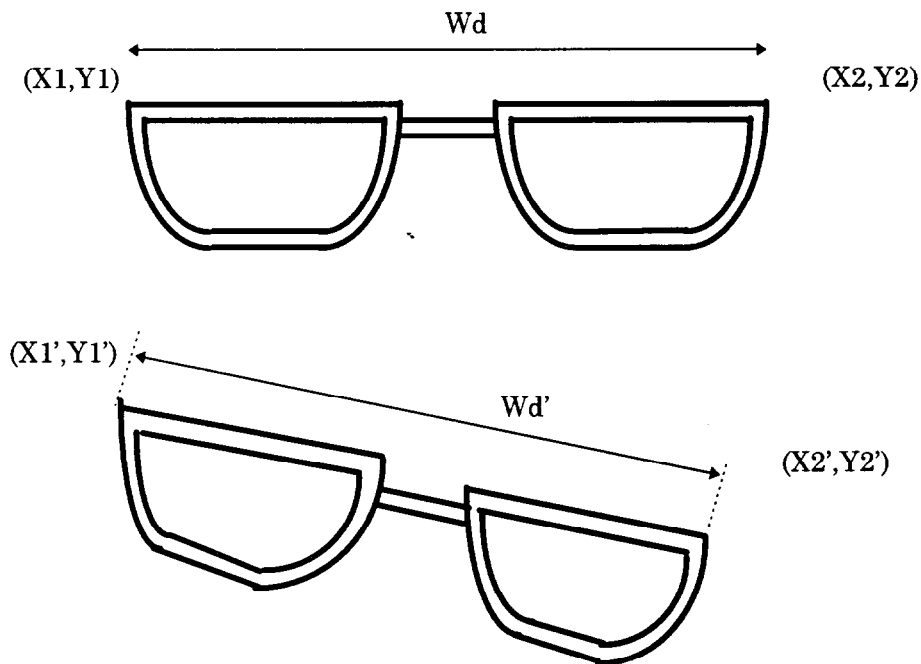


図 5.16 眼鏡フレームの左右端点の位置

(a) 上：初期画像の配置

(b) 下：回転のあるフレームの配置

Position of Left and Right Edges of Eye-glass Frame.

(a) Top :The case for the Initial Frame.

(b) Bottom: The case with Rotating Motion.

得られていることから④の手法で実験を行った。 $(X1',Y1'),(X2',Y2')$ を図 5.16 の2点の座標とすると、傾きは $(Y2'-Y1')/(X2'-X1')$ となる。但し、傾きは通常最大でも 30~40 度であるとみなせるので、90 度になることは無いものとする。抽出された眼鏡フレーム画像をこの傾き値により、眼鏡フレームに向かって左端を中心に回転させる。眼鏡フレームの左端は、顔画像ワイヤフレームモデル上に設定された眼鏡フレーム左端の頂点位置に固定されている。また $(X1',Y1'),(X2',Y2')$ の距離により眼鏡フレームのサイズのスケージングがなされる。これにより右端も顔ワイヤフレームモデルの眼鏡フレームの右端の位置に配置されていることになる。図 5.17(a)に動きパラメータとして3次元回転を含んでいる画像 Claire を使用した場合の処理結果を示す。

5.3.3 消え残りの評価

図 5.17(b)に、表 5.4 のB方式に相当する別系統で入力された眼鏡を上書きした例を示す。但し、これはB方式の様子を調べるためのものであるため、3次元モデル化や3次元処理は省略してある。照明条件が異なるため、眼鏡フレームに違和感がある。またフレーム境界に対して特には処理を行っていないため、境界部分で顔画像との親和性が悪くなっている。一方、図 5.14(c)は初期段階で肌色で上書きした眼鏡フレームの痕跡が残っているが、眼鏡フレームを上書きした図 5.14(d)では、視覚上の感度は低下していると考えられる。

図 5.18 に眼鏡部分を横切る線上の輝度レベルの変化を示す。眼鏡フレームによるレベル変化は最大で 54(13dB、但し最大振幅を 255 としている)、消え残りと思われる部分の変化は 6~14 で平均 10.75(28dB)である。信号としては存在するが、上書きした眼鏡に着目した場合、消え残りは視覚感度的には十分小さくなると考えられる。ある正弦波の場合、コントラストが5倍近いと目立たなくなり、10倍近くあればほとんど気がつかなくなるという研究がある[83]。視覚的には、このように消え残り成分があっても主たる信号との差が一定以上あれば主信号がより強く着目され、消え残りは実際よ

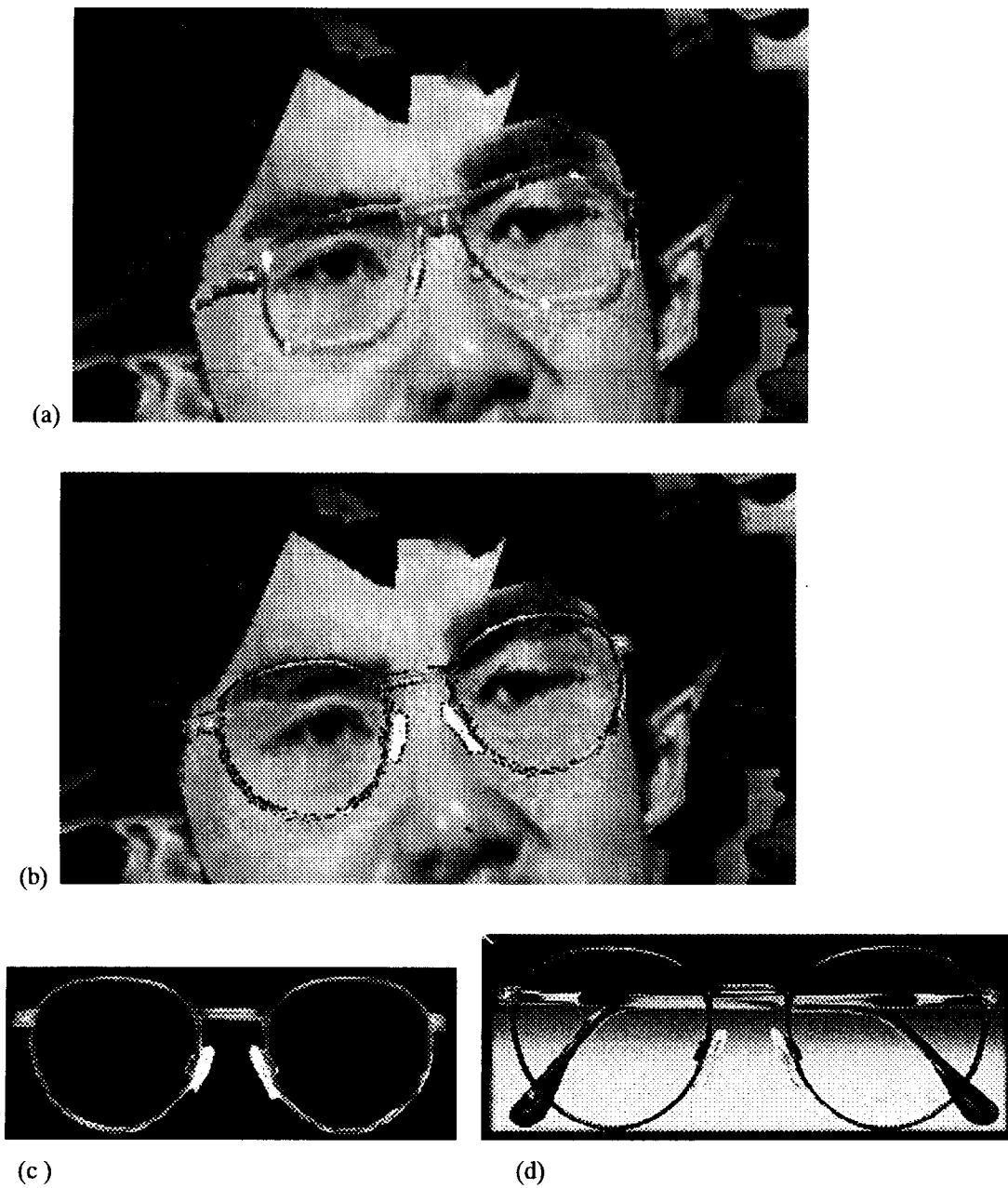


図 5.17 大きい回転のある場合の眼鏡画像処理(claire 第90フレーム)

- (a) 提案する方式A眼鏡画像処理による合成画像
- (b) 別系統入力の眼鏡画像を用いた方式Bの合成画像
- (c) (b)で用いた眼鏡フレーム画像
- (d) (c)を作成するもとになった眼鏡画像

Eye glass processing for large rotation image (Claire, the 90th frame)

- (a) Synthesized image by the proposed method A.
- (b) Synthesized image by the method B with separated input eye glass image.
- (c) Eye glass image for synthesizing (b)
- (d) Eye glass image for making (c)

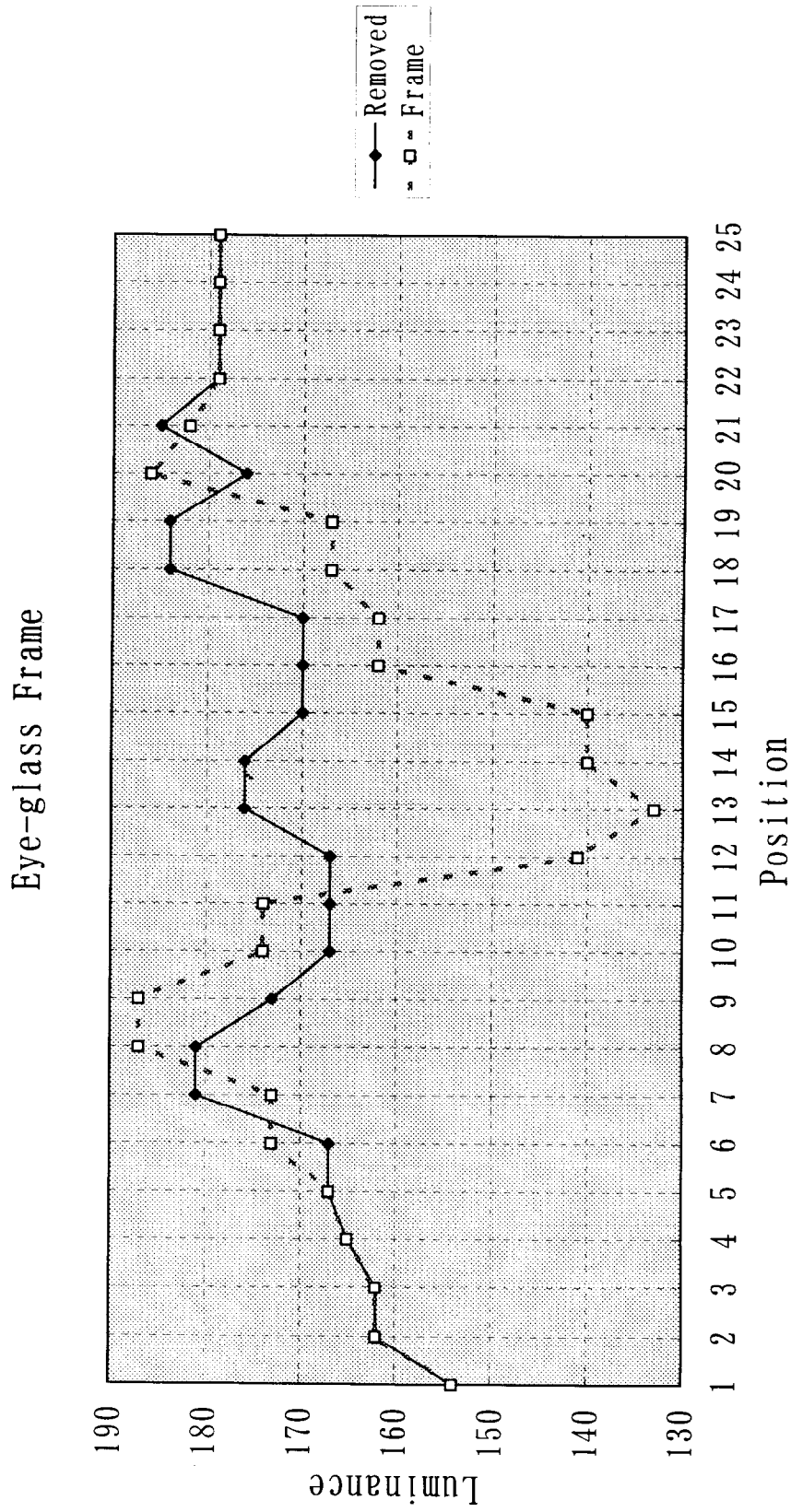


図 5.18 眼鏡フレーム部分を横切る線分上の輝度とその消え残り
Luminance on the Line across Eye-glass Frame and Removed Frame

第5章 ローカル動き検出、表情、眼鏡画像処理

り目立たなくなるという性質がある。そこでテレビゴースト消去実験における主観評価データを用いて評価することにする。

眼鏡処理画像の消え残りは、眼鏡フレーム部の周辺にやや変形したフレーム画像として重畳しているため、テレビ受像機におけるゴーストのような画像となる。テレビゴーストのレベルとその消え残りの主観評価実験が行われている[84][85]が、それによると、事実上気にならないという画質4と気になるレベルである画質3の境界は DU 比[☆]で 24dB となっている。これから、本眼鏡処理においても消え残りレベルの平均がこのレベルに達した時点で眼鏡の消去処理を停止することが妥当と考えられる。ここで、参考として主な D/U 値に対する眼鏡付き画像のゴースト画像を合成し、確認を行った。図 5.17(a)は原画に対し右と下の方向にそれぞれ 25 画素の平行移動を行い、ゴーストとして重ねた画像である。(b)は D/U=18dB、(c)は同 24dB、(d)は同 30dB に対応する。眼鏡部分に関しては、24dB で存在はするがほとんど気にならないレベルであるという基準に合致している。

----- 脚注-----

☆DU 比：希望波 (Desired) とゴースト波 (Undesired) のキャリアレベルでの比で $20\log_{10}(D/U)$ で定義されている。

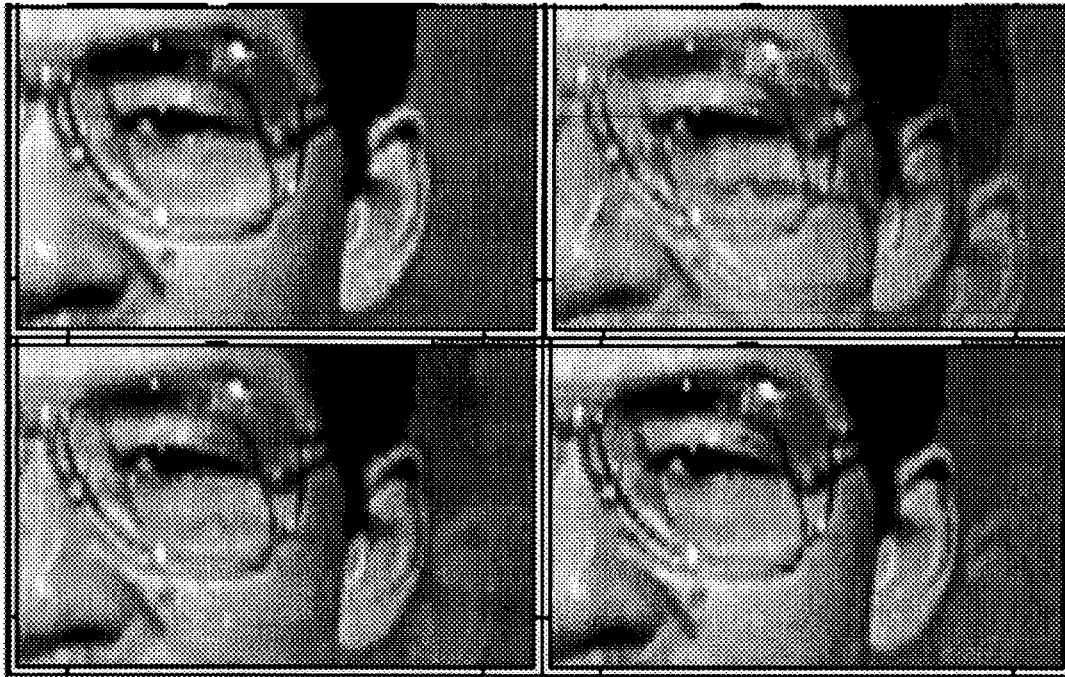


図 5.19 眼鏡部の消え残り評価のためのゴースト画像による比較

(a) 上左：原画像 (13dB) (b)上右：ゴースト画像(18dB)

(c) 下左：ゴースト画像(24dB),(d) 下右：ゴースト画像(30dB)

Comparison of ghosted images for evaluating residual eye glass frame.

(a) Top left: Original image(13dB), (b) Top right: With 18db ghost.

(c) Bottom left : With 24dB ghost, (d) Bottom right: With 30dB ghost.

5.4 むすび

本論文では知的符号化にインタラクティブ操作を導入し、三次元動きパラメータ検出などの問題を解決してきた。

本章では、インタラクティブ知的符号化におけるインタラクティブ操作の曖昧性についての考察に基づき、符号化システムにおけるローカル動き検出以降の部分について、手順の明確化と操作量を評価する検討を行った。口などに対しては、gm→lm 方式が、目の開閉に関しては lm→gm 方式が有利であることを示し、ローカル動き記述順序の操作回数と難易度の比較によりその裏付けを行った。

次に、ワイヤフレームモデル上の顔部品の頂点に階層的な親子関係を定義し、通常の多くの発話シーンで存在する唇の上下動に対し、全点指定の場合に比べ、操作回数と符号量の削減を図る頂点移動方式を開発した。低ビットレート伝送の特徴を生かしたまま、唇両端点に対する0次、0.5次、1次の周期波形をパターン化することにより、閉口やうねるように曲った唇の形状表現におけるインタラクティブ操作量の削減を図った。

開発したインタラクティブ知的符号化の符号化ツールを拡張し、眼鏡処理を用いた符号化を行い、眼鏡フレームに対する肌色による上書き消去処理の操作量の限界が曖昧になることを新たな問題点として指摘した。定量的な判定基準として、テレビゴーストレベルの主観評価実験データを用いて、有限の操作回数で行うための上書き消去処理操作の終了の指標を作成した。具体的には上書き消去処理の不完全性による消え残りが、D/Uで24dB以上になったところで、処理を終了するという操作に対する基準が作成できた。

以上操作量を考慮して、実用的な時間内に符号化操作を行うことのできるローカル動き検出ツールの開発を行った。本方式は別に発表するグローバル動きパラメータ検出ツールと合わせてシステム化し、著者らの他にも技術者が使用し、15種以上の動画系列の符号化が行われている。

第6章 顔動画像の伝送実験と画質評価

6.1 まえがき

上に述べてきた3、4、5章においては、インタラクティブ知的符号化の基本構成、グローバル動き検出方式、ローカル動き検出方式、更には眼鏡画像処理等について、基本構成に対する符号化実験を行い、基本動作の検証を行なって来た。モデルベース符号化や知的符号化における従来の提案や符号化実験に対し、インタラクティブ操作を導入した場合の諸々の問題点を検討し、有効性を確認してきた。本章では、各種応用を想定した場合の伝送実験や画質評価について述べて行く。まず、5章で述べた、グローバル動き検出とローカル動き検出の記述法の違いによる影響を比較実験し、次に方式全体の性能と、モデル化された顔記述を用いた入れ替え実験例、ビット量について述べる。更に、顔動画像伝送をヒューマンエージェントの一形態として応用を想定し、インターネットのホームページでの実験を行っていく。送信側では非実時間環境で、インタラクティブに知的符号化を行い、低ビットレートの符号化データを作成する。次に、インターネットでの伝送の後、復号すなわち再合成を行い、実時間での合成についての評価を行なう。これにより広い意味で知的符号化を一部含んだモデルベース符号化がインタラクティブ知的符号化により、実用化可能であることを示して行く。

6.2 ローカル動きの記述方式による違い

第5章で述べたようにローカル動きの記述方式の違いにより、記述の正確さが異なってくる。図 6.1 は初期フレームに対するワイヤフレームモデルのフィッティング結果である。傾きや回転のない正面画像であるため、目、口ともに正確な位置合わせが行われている。次に、表 5.1 で比較した「初期モデル記述」、すなわち「lm→gm 記述」を行うためには、各フレームでのワイヤフレームモデルのフィッティングが必要であるが、図 6.2 に第100フレームに対するフィッティング結果を示す。顔部品の配置が回転により正面から周辺に移動しているため、輪郭、口周辺にずれが生じている。このワイヤフレームモデルをグローバル動きパラメータを用いて逆変換すると、これらの誤差が変形されて重畳するため、表情などに関連したアクションユニットでの記述に支障をきたす。一方、移動後モデル上での記述(gm→lm)の場合を図 6.3 (a,b)に示

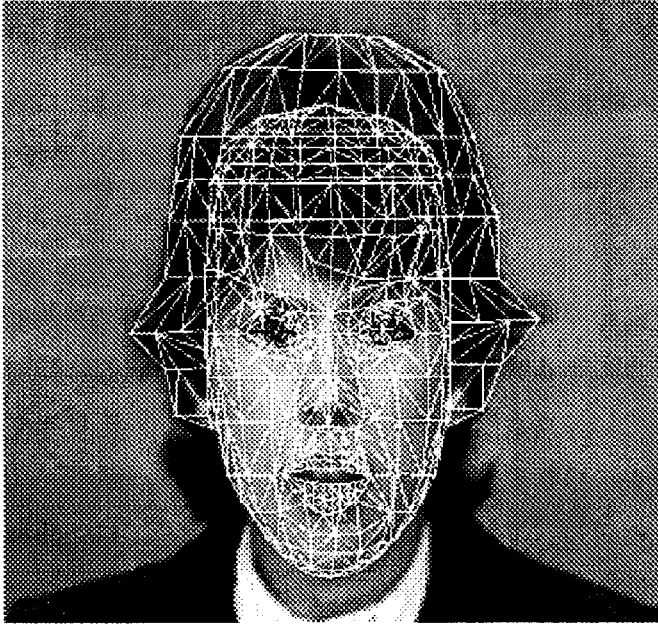


図6.1 第一フレームのワイヤフレームモデル

Wireframe model of the first frame.



図6.2 第100フレームのフィッティング結果

Fitted wireframe model of the 100-th frame.

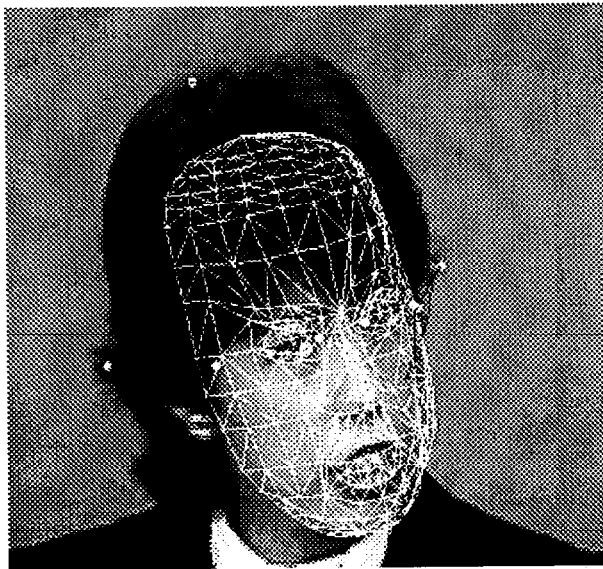


図 6.3 (a) 第 100 フレームの gm 後ワイヤフレーム
Wireframe model after gm transform at the 100-th frame.



図 6.3 (b) 第 100 フレームの gm+lm 後ワイヤフレーム
Wireframe model after gm+lm transform at the 100-th frame.

す。図 6.3 (a)は初期ワイヤフレームモデルにグローバル動きを施した段階の画像で、図 6.3(b)はその後ローカル動きを施した画像である。これを見ると単に口の位置だけでなく、全体のバランスの取れたワイヤフレームモデルが生成されていることが分かる。このワイヤフレームモデルを用いてテクスチャマッピングがなされるため、初期モデル記述(lm →gm 記述)のように、逆変換面上での記述の近似の誤差による新たなずれは生じない。

6.3 再生画像

ローカル動きの記述を移動後記述 (gm→lm 記述) として、単眼カメラ入力の実画像の符号化を行い、画像合成を行った。図 5.1 の符号化ブロック図において、初期画像として図 6.4 (a), (b), (c) に示す顔画像、肩画像、背景画像を用いている。まず、図 6.5 は Claire の復号画像である。図 6.6 は Miss America(MA)の復号画像である。これらの符号化データは初期画像に対してワイヤフレームモデルのフィッティング情報が完備して

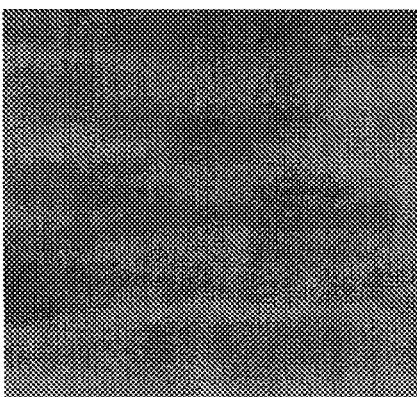
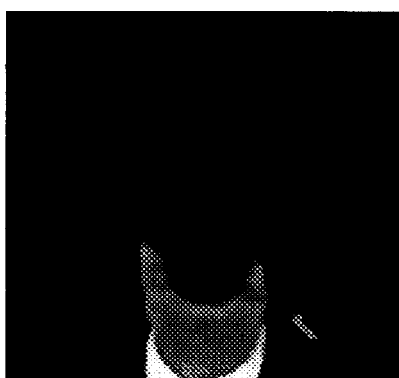


図 6.4(a)顔画像
(第一階層)
Face image
(the first plane).

図 6.4(b)肩画像
(第二階層)
Shoulder image
(the second plane).

図 6.4(c)背景画像
(第三階層)
Background image
(the third plane).

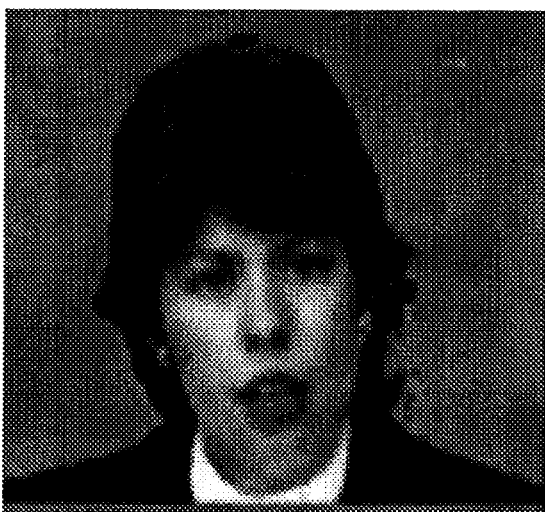


図 6.5 復号画像(Claire)

Decoded image(Claire).

(画像 Claire と Miss America は旧 CCITT(現在の ITU-T)がテレビ会議用符号化方式の国際標準化の際に作成したものを使用した。)

いるため初期画像の入れ替えにより、ある人物に対し別の人物の動きを適用することが容易に行える。図 6.7、図 6.8 は Claire と Miss America を入れ替えた例である。図 6.7 では背景レイヤの画像も入れ替えてある。図 6.4(b)の肩画像は第二階層として、頭部の後ろに配置しているが、口の左右にある顔輪郭上の 2 個の特徴点の平均位置を基に平行移動している。首の部分を上部へ延長して、頭部動きの変動を吸収するようにしている。これにより背景と肩を含めた合成がほぼ自然な形で行えることが分かる。



図 6.6 復号画像(Miss America)

Decoded image.



図 6.7 復号画像

(Texture: Claire, Motion: MA)

Decoded image.

(Texture: Claire, Motion: MA)

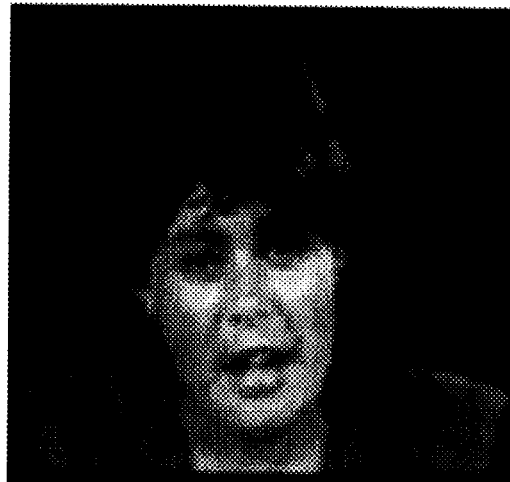


図 6.8 復号画像

(Texture: MA, Motion: Claire)

Decoded image.

(Texture: MA, Motion: Claire)

図 6.7 は細長い顔の Claire が Miss America のワイヤフレームモデルによりやや丸みを帯び、口の形も大きくなり Miss America のように変形していることが分かる。図 6.8 は逆に、Miss America がたて長の顔になり、口は小さくなっているのがわかる。従来、個性を平均化した共通の動きや表情であるアクションユニットを用いることにより個性的な表現が抑圧されていた。このように第5章で述べた「移動後の記述」(gm→lm 記述)によれば共通のアクションユニットによる表現とは異なる実写から入力された個性的な顔の形を正確に記述でき、任意のテクスチャに適用することができる。

ローカル動きの記述を移動後記述(gm→lm 記述)とした通常の符号化を行った場合の復号画像の全部のフレームを図 6.9、図 6.10 に示す。図 6.9 の(a)~(f)は CCITT の Claire の第0フレームから第 118 フレームの画像で、図 6.10 の(a)~(h)は CCITT の Miss America の第0フレームから第 149 フレームの画像である。

以上標準画像を中心に検討したが、他の画像でも実験を行っている。独自にカメラから入力した10数種のあいさつなどの顔画像に対し、同じようにインタラクティブ知的符号化を行い、良好な再生画像を得ている。図 6.11 に合成画像例を示す。

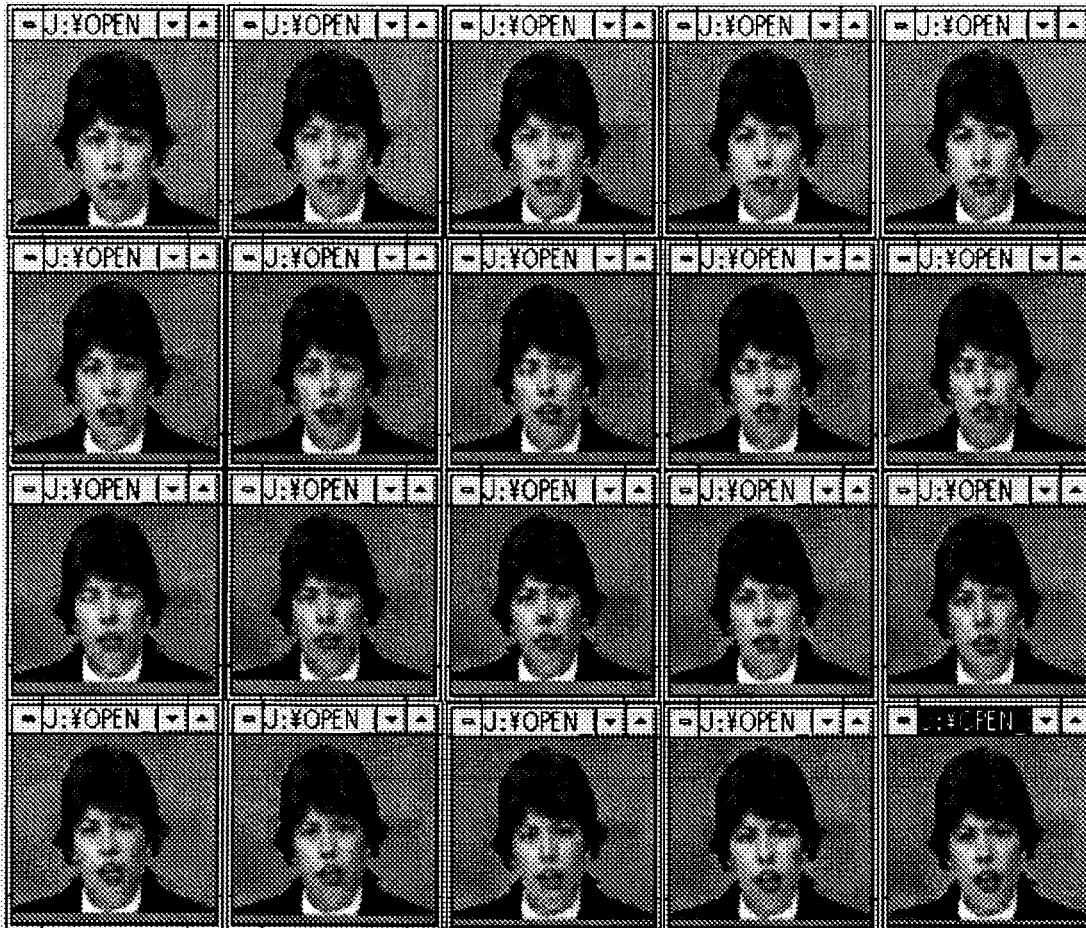


図 6.9 (a) CCITT Claire 復号画像、第 0 フレーム～第 19 フレーム
Decoded images of Claire, from 0-th frame to 19-th frame

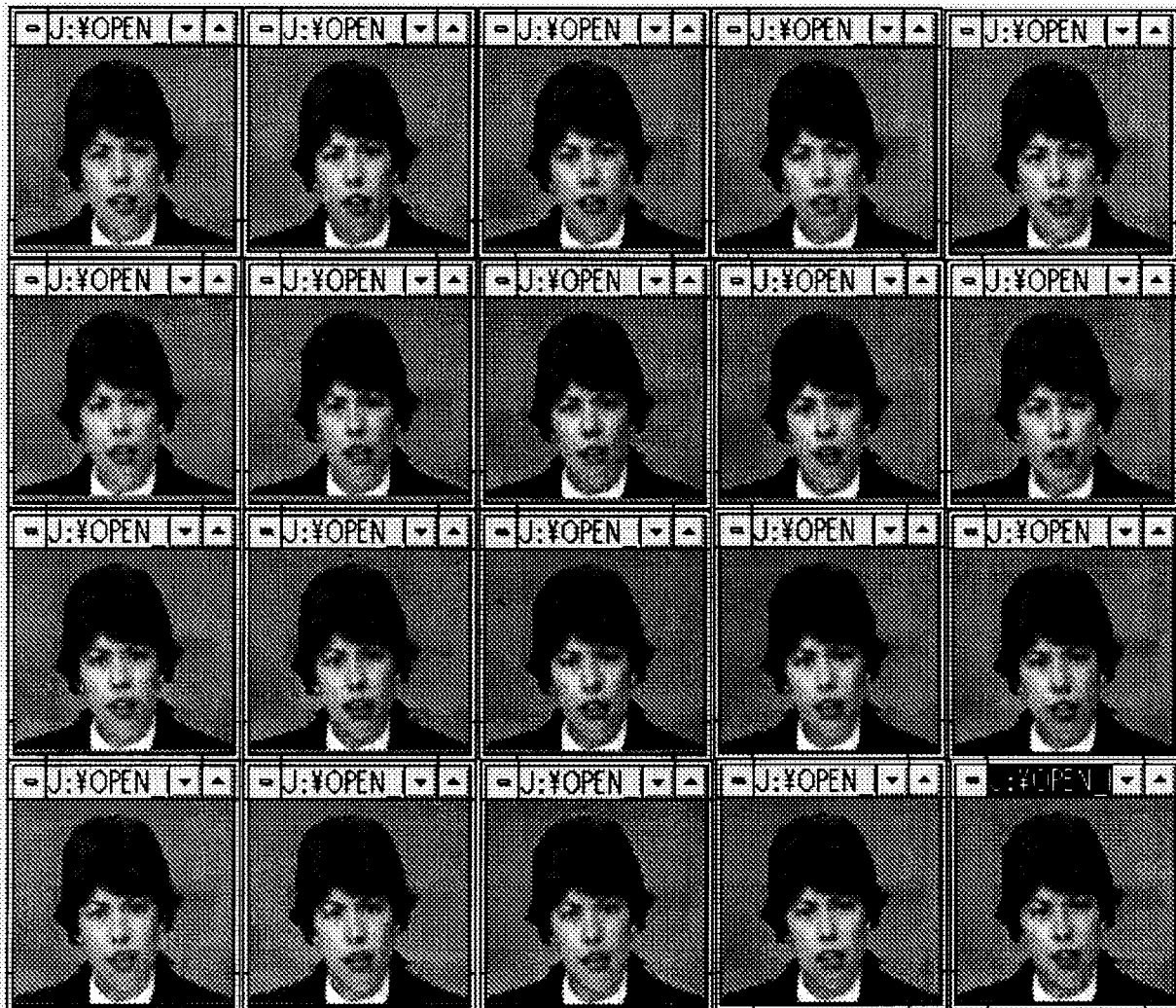


図 6.9 (b) CCITT Claire 復号画像、第 20 フレーム～第 39 フレーム
Decoded images of Claire, from 20-th frame to 39-th frame



図 6.9 (c) CCITT Claire 復号画像、第40フレーム～第59フレーム
Decoded images of Claire, from 40-th frame to 59-th frame

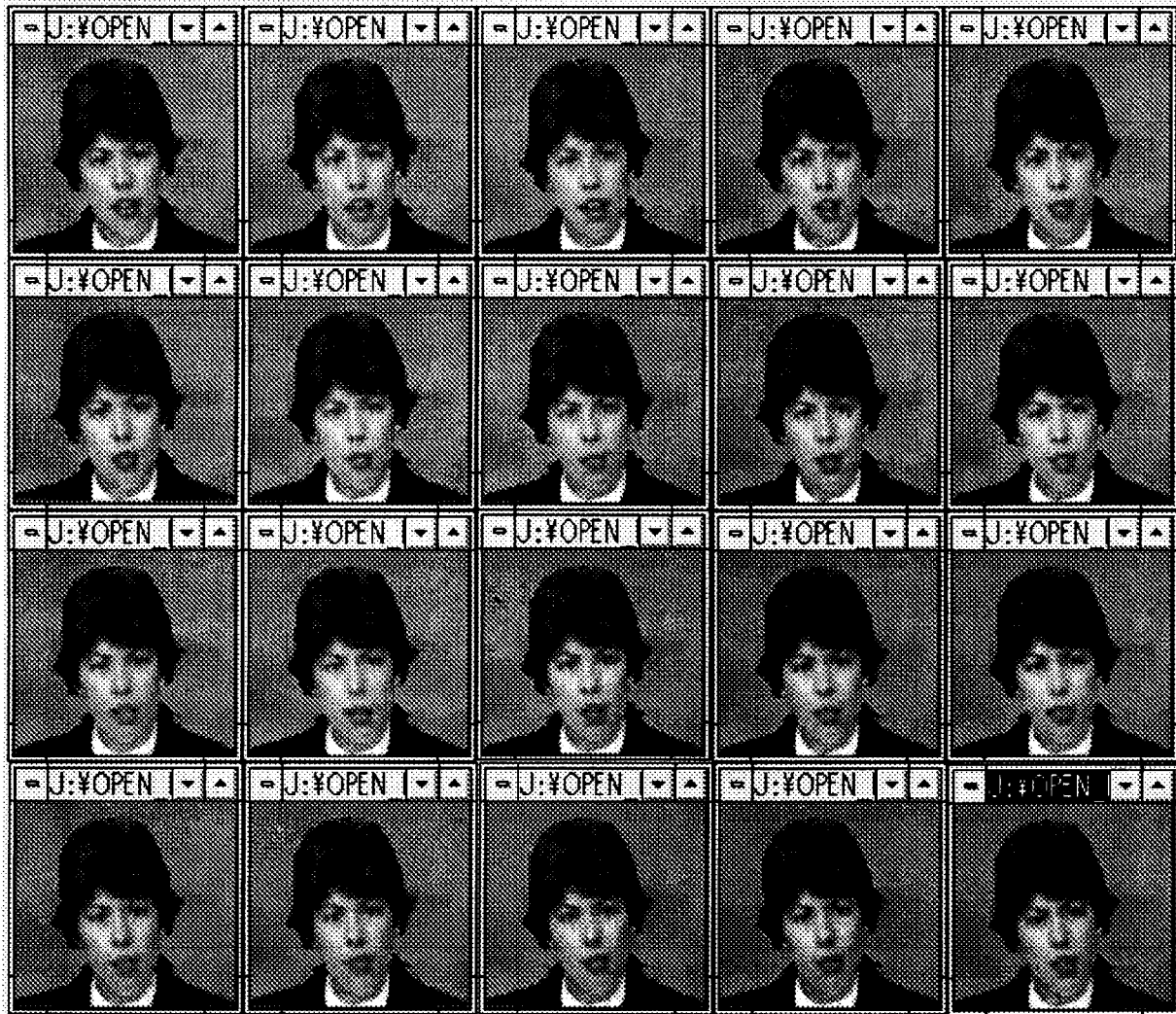


図 6.9 (d) CCITT Claire 復号画像、第 60 フレーム～第 79 フレーム
Decoded images of Claire, from 60-th frame to 79-th frame

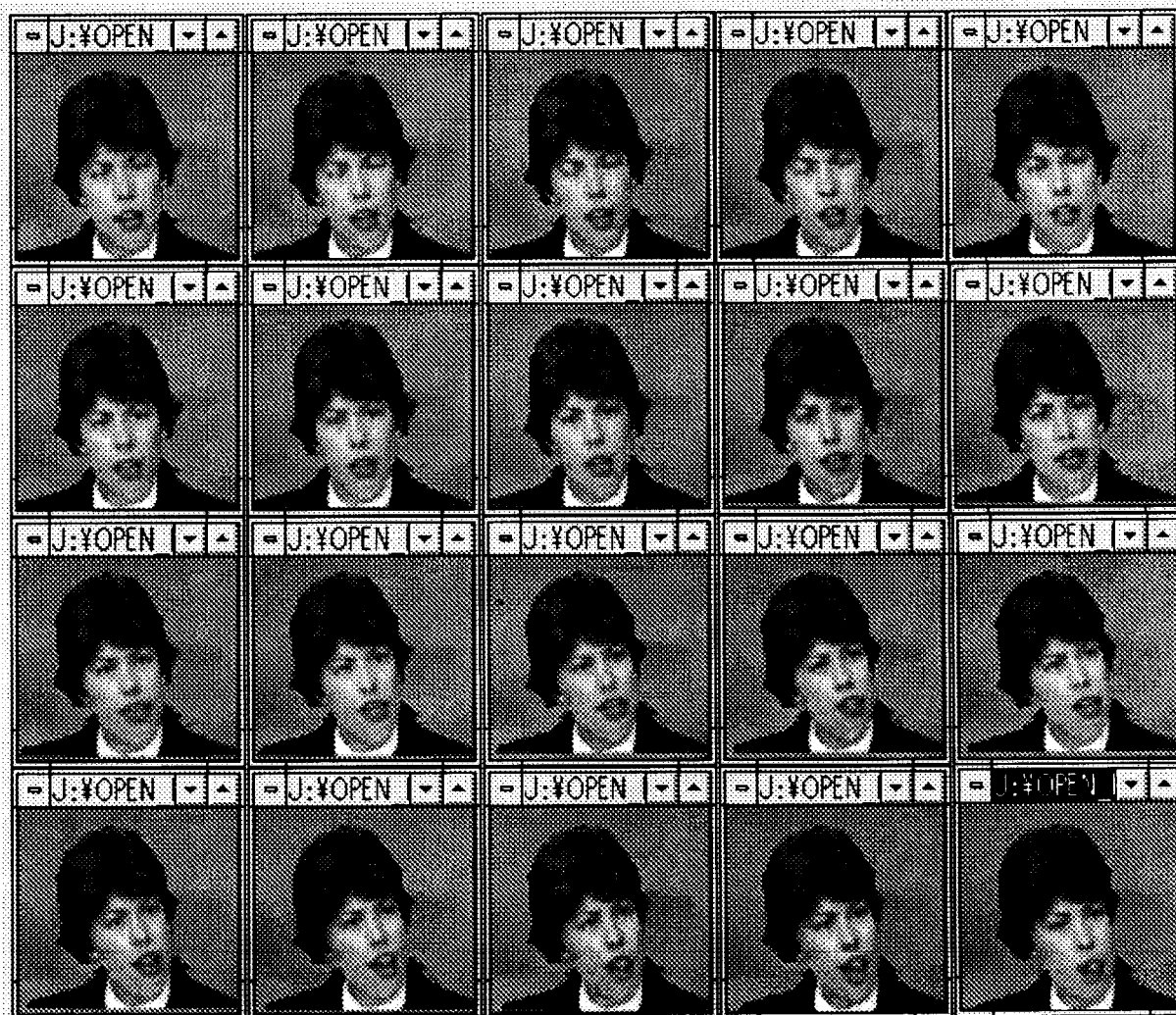


図 6.9 (e) CCITT Claire 復号画像、第 80 フレーム～第 99 フレーム
Decoded images of Claire, from 80-th frame to 99-th frame

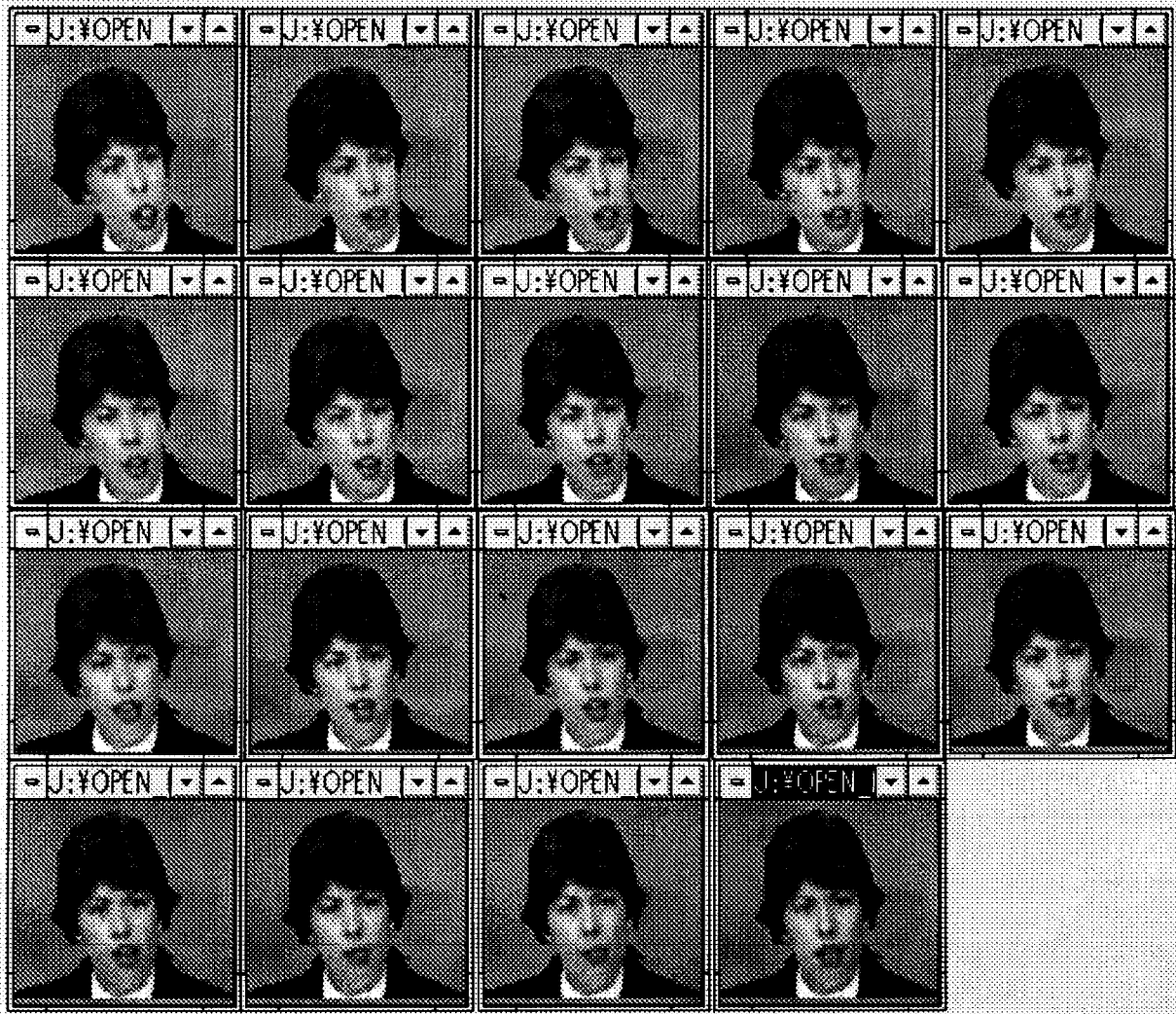


図 6.9 (f) CCITT Claire 復号画像、第 100 フレーム～第 119 フレーム
Decoded images of Claire, from 100-th frame to 119-th frame



図 6.10 (a) CCITT Miss America 復号画像、第0フレーム～第19フレーム
Decoded images of Miss America, from 0-th frame to 19-th frame

第6章 顔動画像の伝送実験と画質評価



図 6.10 (b) CCITT Miss America 復号画像、第0フレーム～第19フレーム
Decoded images of Miss America, from 0-th frame to 19-th frame



図 6.10 (c) CCITT Miss America 復号画像、第 40 フレーム～第 59 フレーム
Decoded images of Miss America, from 40-th frame to 59-th frame



図 6.10 (d) CCITT Miss America 復号画像、第 60 フレーム～第 79 フレーム
Decoded images of Miss America, from 60-th frame to 79-th frame



図 6.10 (e) CCITT Miss America 復号画像、第 80 フレーム～第 99 フレーム
Decoded images of Miss America, from 80-th frame to 99-th frame



図 6.10 (f) CCITT Miss America 復号画像、第 100 フレーム～第 119 フレーム
Decoded images of Miss America, from 100-th frame to 119-th frame



図 6.10 (g) CCITT Miss America 復号画像、第 120 フレーム～第 139 フレーム
Decoded images of Miss America, from 120-th frame to 139-th frame



図 6.10 (h) CCITT Miss America 復号画像、第 140 フレーム～第 148 フレーム
Decoded images of Miss America, from 140-th frame to 148-th frame



図 6.11 他の合成画像例 (モデル: 林直美)
Other synthesis images(Naomi Hayashi)

6.4 ヒューマンエージェントを目指した顔動画像伝送

顔動画像のモデルベース符号化や知的符号化はテレビ電話のような実時間応用としても検討されて来た。知的通信の応用の究極的な検討例として3次元の臨場感通信がある[86]。しかしながら、送信側における画像理解ないしは画像認識などの各種処理を実時間に全自動で行なうことは、現時点ではまだ困難性が高い。ここでは、モデルベース符号化やある種の部分的ではあるが広い意味での知的符号化を現状で実用化するため、送信側の処理を非実時間でインタラクティブに行い、符号化システムを構築して来た。インタラクティブ知的符号化では非実時間環境での応用に限定し、符号化は予めオペレータにより、オフラインで行なわれ、符号化結果は蓄積されている。一方、復号側は原則的に実時間で処理を行なうことをめざしている。第3章での基本検討では、現状のパーソナルコンピュータを用い、インターネットなどの狭帯域通信路を介して顔動画像の伝送と再合成表示を行なうことを目標として、方式開発を行なって来た。このような応用の実用性を検証するため、インターネットに接続したホームページに符号化データを格納し、伝送実験を行った。数秒の顔動画像のデータを伝送する実験であるが、このような伝送と実時間表示が可能となれば、ヒューマンエージェントにおける顔動画像を用いたエージェント機能が実現出来るようになることになる。従って、ホームページでの顔の実動画像を用いた案内や紹介、説明などはもちろんのこと、予め顔動画像の圧縮系列を作成しておくことが可能な応用に広く適用可能となる。

以下、通信・放送機構、映像分析プロジェクトのホームページを用いて行った顔動画像の伝送実験について、具体的に述べて行く。

6.4.1 顔動画像の伝送実験の構成

図6.12は通信・放送機構、映像分析プロジェクトのホームページの顔動画像伝送実験に関する部分のファイル構成である。映像分析プロジェクトのページから下記のような4階層のファイル構造になっている。

① <http://vega.tao.or.jp/mvp>

② <http://vega.tao.or.jp/mvp/members.html>

③ <http://vega.tao.or.jp/mvp/ohzeki/index.html>

④ <http://vega.tao.or.jp/mvp/ohzeki/tool/index3.html>

第4の階層に符号化データに復号・画像合成プログラムと説明書一式が zip ファイルに圧縮され、PC上で自己解凍する、

MBC34CZ.EXE (244,839 バイト)

MBC3MZ.EXE (263,551 バイト)

等の形で格納されている。HTMLブラウザ上でこれらの名称をクリックすれば、ダウンロードされる。受信側でこのファイルを実行すると、図 6.13 に示すような表示があらわれる。

すなわち、まず WinZip Self Extractor の表示がなされ、解凍するディレクトリの確認がなされる。デフォルトで

C:¥MBC¥MBC2

が表示されているので、必要により適宜変更を行なう。この場合は、

G:¥WORK

に変更し、Unzip をクリックする。解凍処理がなされ、12ファイルが生成したことを確認する。OKボタンをクリックすると、以後、MS-DOS プロンプトにて、JPEGにて圧縮された初期画像3枚を.PPMのファイルに復号する。尚、JPEG復号プログラムも最初のファイルに含まれている。この処理として、下記のようなコマンド系列が解凍後に実行されるようになっている。

```
djpeg.exe -ppm T-IM.JPG T-BGIM.JPG T-BODY.JPG | mbc2.exe | exit
```

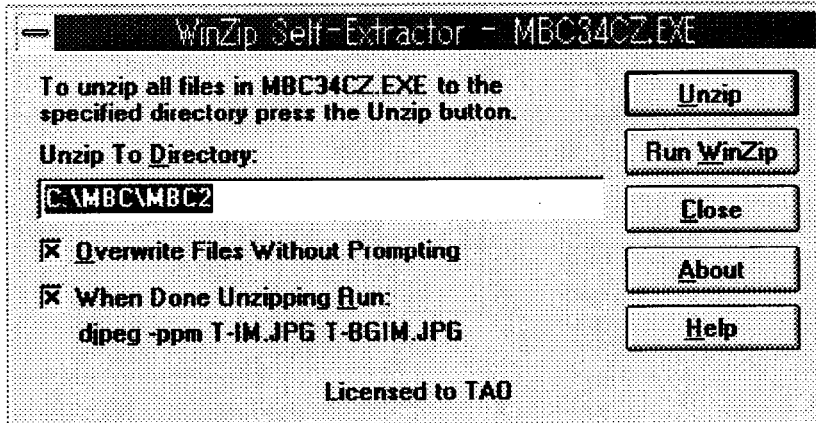
この後、合成処理に入り、必要フレームを合成後し、動画像として表示する。

表 6.1 に実験に使用した符号化ソフトウェアツールを、表 6.2 に復号実験に用いたコンピュータの仕様を示す。

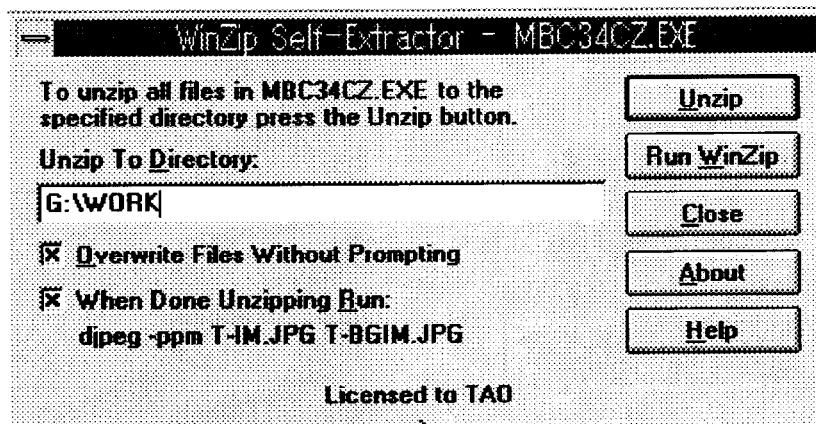


図 6.12 通信・放送機構映像分析プロジェクトの顔動画像伝送実験のファイル

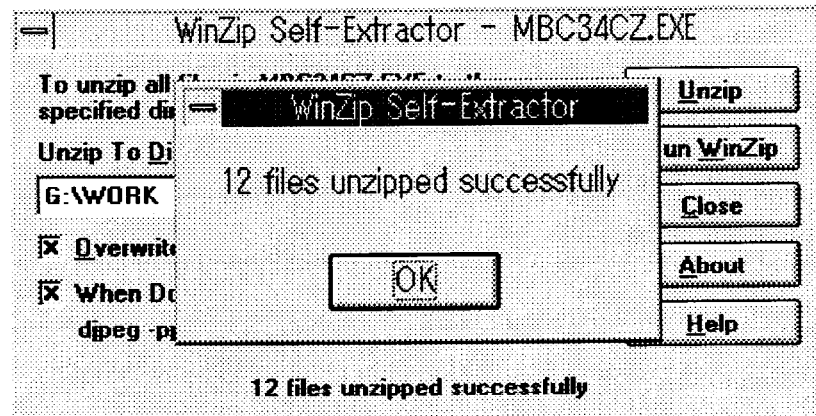
Files used in facial image sequence transmission experiment in Media Vision Proj. of TAO



(1) WinZip Self Extractor の表示がなされ、解凍するディレクトリの確認を行なう。

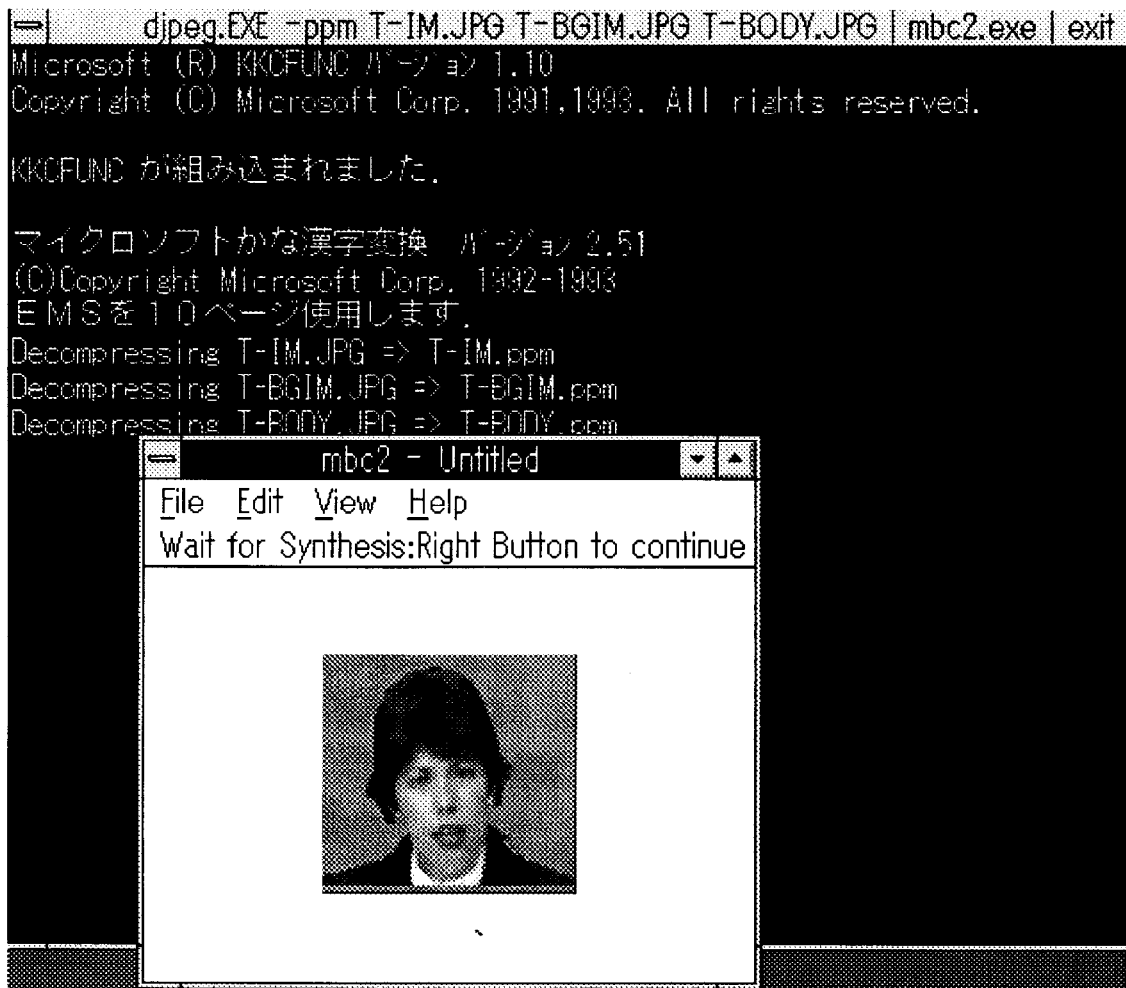


(2) G:¥WORK に変更し、Unzip をクリックする。



(3) 解凍処理がなされ、12ファイルが生成したことを確認する。OKボタンをクリックすると、以後、MS-DOS プロンプトにて、処理がすすむ。

図 6.13(a) ダウンロード後の実行時の表示—その1
Downloading procedure and running(1)



JPEGにて圧縮された初期画像3枚を.PPMのファイルに復号後、
合成処理に入り、必要フレームを合成後し、動画像として表示する。

図 6.13(b) ダウンロード後の実行時の表示—その2

Downloading procedure and running(2)

表 6.1 顔動画像伝送実験使用符号化ソフトウェアツール

Software tool used for encoder of facial image
transmission experiment

項目	ソフトウェア名称
OS	WindowsNT3.51
符号化プログラム開発ツール	MS Visual Basic 4.0
復号プログラム開発ツール	MS Visual C++ 4.0
添付プログラム	djpeg.exe
	winzip self extractor

表 6.2 顔動画像伝送実験復号実験機仕様

Specifications of computer used in facial image
transmission experiment

項目	機種 1	機種 2
構成 CUP	PC/AT Pentium90(DELL)	PC/AT PentiumPro200 (Proside)
OS	WindowsNT3.51、 WindowsNT95	WindowsNT4.0
Graphic	DELL	Force OpenGL

6.4.2 顔動画像の伝送実験の結果

表 6.3 は CCITT の Claire と Miss America の符号化データのファイルサイズである。いずれも上記のとおり、受信側で再合成するために必要な解凍や JPEG 復号プログラムを含んでいる。表 6.3 において太枠で無い部分は入力と同じ動きファイルを用いた場合の動画像である。太枠で囲んだ部分は入力動画像とは異なる動きファイルを適用した場合である。

表 6.3 符号化データのファイルサイズ

File size of coded data

画像サイズ	動きファイル	CCITT Claire	CCITT Miss America
128x120	Claire	① 244,839 B	270,794 B
	Miss America	② 291,559 B	263,551 B
256x240	Miss America		263,551 B

このように、本インタラクティブ符号化システムにおいては、符号化データが単なる圧縮されたビット系列というだけでなく、いくつかの機能を持った部分集合にいわば、構造化されているため、復号を介す前にそれらを入替えることが可能である。この動作の様子を図 6.12 に示す。初期画像へのワイヤフレームモデルのフィッティングにおいては、標準のワイヤフレームモデル

wfm00

を符号化対象画像の上に重ねていく。今、Claire に対するフィッティングを行なうと、wfm00 は wfmC0 と変形される。この初期画像の対応関係をもとに、以後の符号化対象フレームの合成が行なわれる。別の人物 Z の顔テクスチャを Claire の顔形状と動きに合わせて合成しようとする時は、Claire のワイヤフレームモデルの各 3 角パッチと人物 Z のワイヤフレームモデルの 3 角パッチの対応がなされる必要がある。具体的には人物 Z の初期画像の他に、claire の初期画像にフィッティングしたワイヤフレームモデルの他に、人物 Z の初期画像にフィッティングしたワイヤフレームモデルが必要となる。表 6.4(a)は表 6.4 の CCITT の Claire の縦方向に続く 2 つのファイル①、②の内容を示したものである。

表 6.4 通常の合成と動きファイルを入替えた合成のファイルの内訳

Detailed contents for own motion file and other motion file

	①テクスチャ: Claire 動き: Claire	②テクスチャ: claire 動き: Miss America
初期画像	T-IM.PPM, T-BODY.PPM, T-BGIM.PPM(Claire)	T-IM.PPM, T-BODY.PPM, T-BGIM.PPM, (MA)
ワイヤフレーム モデル	t-wfm.dat	T-WFM.DAT, T-WFM2.DAT
グローバル動き パラメータ	T-GLOBA.DAT	T-GLOBA.DAT, T-GLOBA2.DAT
ローカル動きパ ラメータ	T-LOCAL.DAT	T-LOCAL.DAT, T-LOCAL2.DAT
その他、共通フ ァイル	mbc2.exe, t-ldata.dat, readme5e.txt, jpeg6e.txt,	t-genet.dat, djpeg.exe,

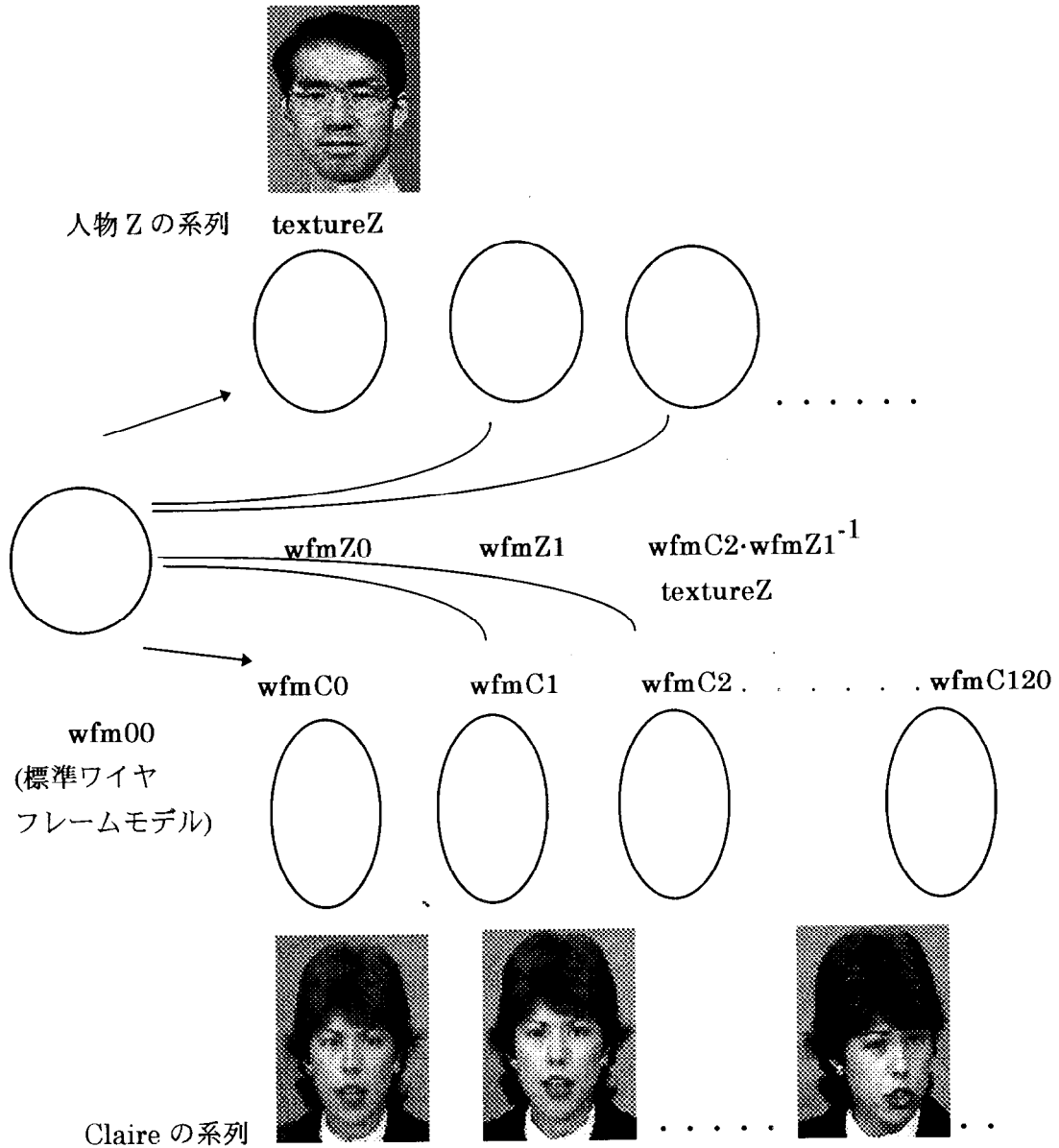


図 6.14 2つのワイヤフレームモデルと動きファイルの関係

Two wireframe models and motion file relation

表 6.5 に合成時間の比較を示す。用いた計算機の機種は表 6.2 に示した 2 機種である。機種 1 は通常の市販の PC (DELL XMT590, Pentium90)をそのまま用いている。1 駒の合成処理に約 2 秒要し、全フレーム合成後は実時間での表示が可能となる。一方、機種 2 は同じく通常の市販の PC であるが、表 6.2 に示すように CPU に PentiumPro200 を使用し、グラフィックボードに OpenGL 対応のものを用いている。これにより、120 駒、約 4 秒の系列が 4～5 秒で再生出来、通常の PC でも実時間での合成が可能であることが示された。

表 6.5 に合成時間の比較

Comparison of synthesis time

項目	機種 1	機種 2
Claire	120 秒	4～5 秒

6.4.3 ヒューマンエージェントとしての顔動画像符号化実験

符号化実験の画像として、主に CCITT のテレビ電話・会議標準化で使用された画像である Claire, Miss America を中心にを用いてきた。このような標準画像以外にも、独自にモデルの撮影を行い、各種のシーンの符号化実験を行った。6.2 の最後でも用いたが、モデル（林直美）に対し、日常の挨拶や説明、劇のせりふなどの演技をしてもらい、合計 1 時間以上にわたる数十種類のシーンを収録した。この中から、表 6.6 に示す 15 種のシーンについて、インタラクティブ知的符号化を行い、再生画像の確認を行った。このインタラクティブな符号化操作は第 3 者にも行ってもらい、一定の基準で操作が可能であることを確かめた。図 6.15 に初期画像を、図 6.16、図 6.17 に挨拶のシーンを符号化したものを復号した再生画像を示す。

表 6.6 ヒューマンエージェント応用のための実験シーン

Test Sequence for application of human agent

おはようございます
こんにちは
こんばんわ
さようなら
はい
いいえ
私の心は二つにひきさかれております。
そんなことをおしゃってもやきもちやきは聞きやしませんよ。
男のかたって本当におかしいのねえ！
おやすみなさい。
はじめまして。どうぞよろしく申し上げます。
私たち、結婚しました。
どうも、ありがとうございます。
あー、びっくりした。

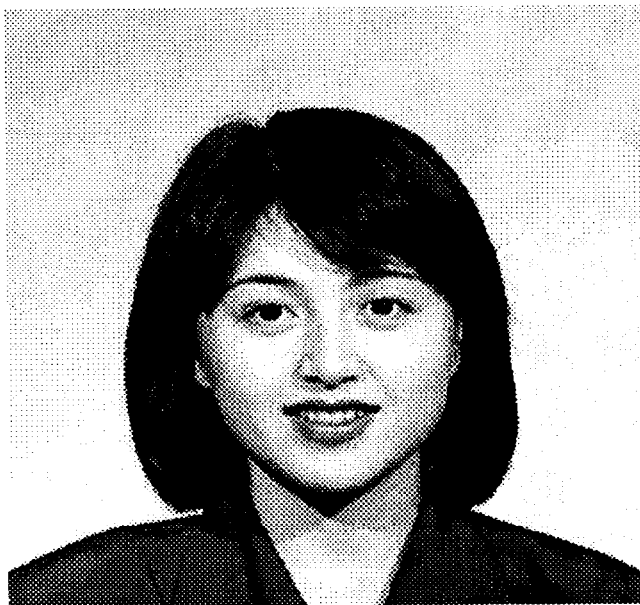


図 6.15 独自画像（林直美）の初期フレーム画像

Initial frame image of non-standard test sequence 'Naomi'



図 6.16 系列 1（おはようございます）の復号画像

Decoded image sequence of Sequence 1('Ohayou-gozaimasu')



図 6.17(a) 系列 2 (こんにちは) の復号画像

Decoded image sequence of Sequence 2('Kon-nichiwa')



図 6.17(b) 系列 2 (こんにちは) の復号画像
Decoded image sequence of Sequence 2('Kon-nichiwa')



図 6.17(c) 系列 2 (こんにちは) の復号画像
Decoded image sequence of Sequence 2('Kon-nichiwa')

6.5 主観評価実験

再生動画像に対して、主観評価実験を行った。本論文で提案している方式と、超低ビットレート符号化で標準方式になっている、H.263[87]方式の比較を行った。標準方式に合わせテスト画像は、Claire と Miss America を用いた。結果を図 6.18 に示すが、本方式によれば、Claire、Miss America の実際のビットレートがそれぞれ 54kbps、46kbps であるが、相対的な比較結果から求まる評価の平均値は H.263 方式よりも良くなっている。この主観評価結果から、総合的に見ても、インタラクティブ的符号化方式の有効性が確認された。評価者は12人で、技術者以外の人や女性事務員らを含んでいる。H.263方式のビットレートと画質を基準にして、相対比較により本方式の画質の位置するビットレートの位置にプロットしてある。

6.6 ビット量の評価

動きパラメータの伝送フォーマットの例を図 6.19 に示す。また、ビット量の内訳を表 6.7 に示す。復号画像を合成するためには、これら全部のファイルが必要になるが、一旦受信した後は、デコーダ等のデータに依存しないファイルについては再送信の必要はない。データに依存しないファイルは、復号（合成）プログラム mbc2.exe、ローカル動き記述時の頂点間の連動規則のファイル ldata、JPEG 形式で受信する3層の画像を復号する djpeg である。2回目以降必要となるのは、初期画となる bgim、body、im の3層の画面データと、顔のワイヤフレームモデルのフィッティング情報 wfm、グローバル動きファイル globa、ローカル動きファイル local である。表中、ZIP 圧縮後の欄は、各ファイルの原サイズを圧縮したもので、画像の3ファイルは、JPEG 後のファイルを圧縮したものである。整数化の欄の3個は図 6.19 に示した符号系列として8ビットで表現したもので、ハフマン符号等の可変長符号化は適用していない。Claire の場合これらのサイズの合計は、27.1 kB である。このうち初期画像を除いた動き情報は、3534バイト（28,272ビット）である。ここで符号化した Claire の画像は4秒分であるが、ブロック歪みや駒落としの全くない画像が再生できていることを考えると、効率が高いことが分かる。

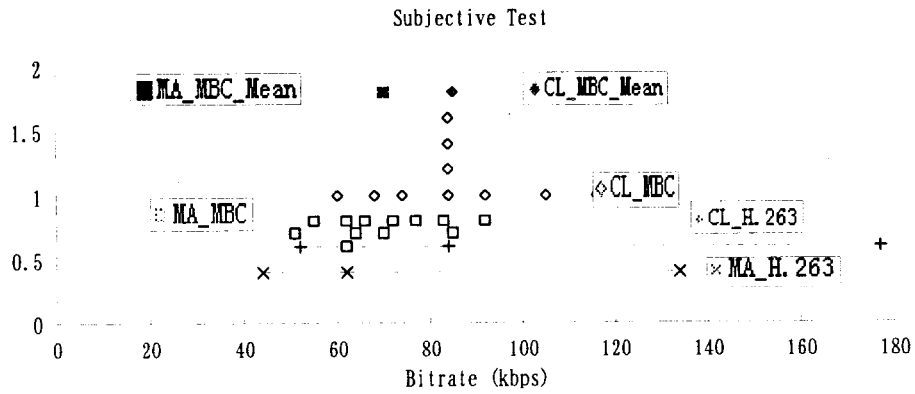


図 6.18 主観評価結果

Subjective Test Result

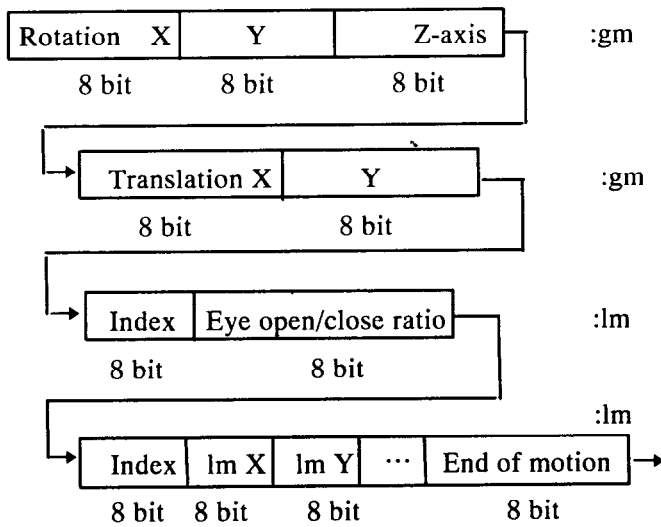


図 6.19 動きパラメータの符号系列の例

An example of code sequence of motion parameters.

表 6.7 ビット量の内訳(Claire)

Detailed amount of bit assignment (Claire).

ファイル名	原サイズ	J P E G 後	ZIP 圧縮後	整数化
mbc2.exe	228kB		111kB	
bgim.ppm	737k	5.8k	<u>4.1k</u>	
body.ppm	737k	10 k	<u>7.7k</u>	
im.ppm	737k	13.5k	<u>11.6k</u>	
genet.dat	21k		7.3k	
globa.dat	7.7k		3.3k	<u>708B</u>
ldata.dat	29k		3.3k	
local.dat	9.5k		2.5k	<u>2826B</u>
wfm.dat	32k		11.1k	<u>168B</u>
djpeg	108k		61k	

6.7 むすび

本章では第 3, 4, 5 章で提案して来たインタラクティブ知的符号化を、種々の応用を念頭においた実用化実験を行った。

ローカル動きパラメータとグローバル動きパラメータの記述順序の違いによる影響を画質の面から検討した。知的符号化で従来より主に行なわれていた $lm \rightarrow gm$ 記述は対象とする顔動画のアクションユニットが既知の場合は有効である。現状ではこのアクションユニットが自動でも手動でも完全には検出できないことを考えると、 $lm \rightarrow gm$ 記述は仮想的な顔動画系列を生成する場合には有効であるが、入力の実画像に忠実な画像合成を行なう場合は、正確な記述ができなくなる。実際の符号化実験を行い、この差異について検証を行った。

次に、初期画像として、顔動画の頭部部分の他に肩の動きを別に記述するための首から下の画像と背景画像との 3 枚を用いる符号化実験を行った。CCITT 標準画像の Claire 約 4 秒 (120 フレーム)、Miss America 約 5 秒 (150 フレーム) の符号化を行い、良好な結果を得た。また、ワイヤフレームモデルと動きパラメータ等に構造化された形で記述された符号化データであるため、復号することなく顔のテクスチャを入替えたり、動きを入替えたりすることが出来るが、その検証実験も行った。また、このような標準画像のみでなく、独自に入力した多数の顔動画に対して符号化を行い、提案して来た符号化システムが汎用的に使用可能であることを検証した。

第6章 顔動画像の伝送実験と画質評価

エージェント研究の画像面の一分野であるヒューマンエージェントにとっては顔動画像の転送、表示が重要な技術となっている。インターネットのホームページのようなハイパーメディアのデータベースにインタラクティブ知的符号化結果を格納し、転送、再生、表示の実験を行った。通常のパーソナルコンピュータでも合成処理に多少時間をかければ、その後はリアルタイムに表示が可能となること、更に OpenGL のグラフィックボードを備えた場合はインターネット回線での遅延を除いて、ほぼ実時間に再生、表示が可能であることが確認された。これらの実験から、通常のパーソナルコンピュータでも実画像に対する実時間の復号、表示が可能で、インタラクティブ知的符号化がヒューマンエージェントを広く実用化していくための基本技術として有効であることが示された。

汎用性に関しては、独自に入力した画像から、15 種以上のものについて符号化を行い、主に顔動画像によるヒューマンエージェントを想定した応用などに使用可能である見通しを得た。また、これらの顔動画像系列の符号化は、筆者以外の技術者が操作して行ったものであり、開発したインタラクティブ知的符号化システムが個別に特殊なようなものでなく、一般性のあるものであることも確かめて来た。

また、再生画像の品質に関しては、従来の低ビットレート波形符号化の代表的存在である H.263 を用い比較実験を行うた。画質とビットレートの比較で、提案して来たインタラクティブ知的符号化がほぼ同一かやや優れた性能を示していることが確かめられた。

今後は、符号化データのハフマン符号化、より多くの顔画像でのワイヤフレームモデルの入れ替え実験などを行う必要がある。

第7章 結論

画像の高能率符号化研究が波形符号化から構造的符号化へ展開している中で、構造的モデルを取り入れた顔動画像の知的符号化を取り上げ、インタラクティブ操作との関連についての検討を行い、インタラクティブ知的符号化の提案を行なった。構造的符号化におけるインタラクティブ操作の意義として、単に人間が補助を行なうという曖昧な方式を放置することなく、工学的に評価可能な規範を提案して来た。本研究により、人間のインタラクティブ操作の程度や数値的基準などの規範のあるインタラクティブ知的符号化システムを構築し、符号化実験によりその有効性を確認した。また、インターネットを用いた顔動画像伝送実験により編集・加工等に関する実験も行なった。知的符号化では初期画像を高能率符号化する必要があることからわかるように、従来の波形符号化も重要な要素技術の一つである。そこで、本研究のはじめの部分では、適応KL変換符号化方式の提案を行い、DCT方式との比較を行なった。

研究成果を要約し、更に今後の課題・展望について述べる。

第1章「序論」では、本研究の背景と目的を述べた。

本研究の背景として、

- <1> 画像の高能率符号化は波形符号化を中心に国際標準化が進んでいる。
- <2> 国際標準化では理論的基盤もあるが、実験を中心とした膨大な経験的結果の中から性能の優れた方式が採用されている。
- <3> 次世代符号化として構造的なモデルを用いた先端的な符号化方式の研究がなされているが、画像の認識の問題で完全な自動化処理には至っていない。
- <4> 画像メディアがマルチメディア化し、テレビ電話などの代表的な応用に加え、インターネットなどの非実時間の画像通信をはじめとした多面的な応用の可能性が増えつつあること、などを述べた。

これらの背景の中で、本研究の目的として、まず波形符号化においては、努めて理論的な観点で効率的な手法についての最適化を検討し、理論的な手法で限界を探求して行くアプローチを行なうことを述べた。次に、知的符号化の問題点を再検討し、これまでのようなインタラクティブ操作の消極的な使用とは異なる、積極的な意味での

第7章 結論

導入を提案している。画像通信の多面的な応用の中で、非全自動で非実時間の符号化を検討し、その課題の分析と解決を行なって行くことを述べた。

第2章では、まず、従来方式の調査を行い、2乗誤差最小化を行なうKL変換は古くからあり、いくつかの証明が示されていることを述べた。変換行列を適応的に切替える適応KL変換符号化方式の構想も1970頃に見られるが、KL変換を設計するカテゴリとして明暗の輝度など、KL変換が有効となるような画像の局所的性質とは言い難い分類を行なっているなどの問題があった。本研究ではDCTなどとの比較を明確化するため、平均値分離を行なわない形式のKL変換行列の導出を行い、それを用いた変換を切替える適応KL変換符号化方式を構成した。この際、特異値展開符号化方式(SVD符号化方式)との違いについての解説を行った。提案した適応KL変換符号化方式の変換行列を実画像データを用いて設計し、符号化実験を行った。提案する設計法で求められたKL変換行列から得られる符号化特性はトレーニング画像の違いに依存しない結果が得られ、非定常性に対して汎用なシステムが構成されているものと考えられる。本設計方式の検討において途中の段階で求まる固有値は事実上全て正で求まり、安定な設計が保証できることも示す。DCTと比較してS/Nで約2 dBの性能向上が得られることが示された。

第3章では、1980年代から研究が行なわれてきた、動画像のモデルベースト符号化、知的符号化の研究動向を整理し、新たに「インタラクティブ知的符号化」の基本構成を提案した。これは第3世代から第5世代にまで及ぶ符号化方式であり、インタラクティブ操作を導入することにより、従来の研究の一部が加速されるという効果をもたらすことを述べた。また、インタラクティブ操作の意義について論じ、単に人間が符号化の支援を行なうということだけでは無く、積極的な意義を備えた方式と成るものであることを述べた。これを3つのポイントにして述べる。即ち、

- (a) 画像認識、特徴検出の新たな中間的な段階を設定し、対象の限定を行なう。
- (b) 知的符号化符号化の画像認識部以外の高性能化、高機能化を図る先行的アルゴリズムの開発が行なえる。

——髭、眼鏡画像処理、背景処理など

- (c) 表情の加工・編集など操作者の意図を反映した創造的処理が可能となる。

送受を問わず、画像テクスチャや動きを入替えたり、加工することが可能となる。

第7章 結論

全自動の画像の特徴点検出・追従などの認識的な部分は、インタラクティブ操作によりどのような情報を付加すれば、どのような検出が可能かを調べることにより、新たに全自動と、マニュアルの間にインタラクティブ操作を用いた中間的な処理を作り出すことができる。また、知的符号化が操作量に規範を与えた形で実現できるように成り、従って知的符号化そのものの高機能化や多機能化が図れるようになる。更には構造化された諸データを活用し、画像の編集や加工の処理がインタラクティブ操作の積極的な意味として実現できるようになる。

応用としては、MM-email 用の「意図・符号化・編集機」を想定し、分析合成符号化に人間の操作を対話的に付加するシステムを検討した。具体的には、従来の分析合成符号化システムを統合し、これに、意図を画像に反映させる対話型編集システムを構築することができる。また、従来はテレビ電話のような固定化した代表的な応用に対する検討が中心であったが、画像メディアの多面的な発展に対応して、非全自動で非実時間の処理を行なう符号化方式も重要な応用となってきている。また、従来の知的符号化ではむしろ回避されていたような眼鏡や髭のある顔動画像の符号化処理も可能となって来る。以上、第3章ではインタラクティブ知的符号化の基本構成の提案を行い、構造化や対話型編集器の構想について述べた。

第4章では、インタラクティブ知的符号化の送信側符号化ツールのうち、頭部グローバル動きパラメータ検出にかかわる部分における問題点とその対策について検討した。インタラクティブ操作の有効性について検討を行い、全自動とマニュアルの間に、人間がどのような入力情報を与えれば、どのような中間的な段階が設定出来るかの検討を行った。GUIを用いて比較的入力し易い情報として顔画像上の特徴点を検討した。その結果、顔画像上に12点の特徴点を定め、全自動とマニュアルの中間的段階として2次元から3次元の動きパラメータを検出する課題を設定した。この課題に対し、顔の3次元ワイヤフレームモデル等の初期条件を加え、奥行きとグローバル動きパラメータを求める2段階からなるアルゴリズムを新たに開発した。動きパラメータの検出は、従来2次元観測データから得られる9個の連立2次方程式を解かなければならなかったが、特徴点3個の組合わせから求まる3個の連立2次方程式による解法を導き、より正確な動きパラメータの検出方式が開発できた。線形方程式により、グローバル動きのパラメータ検出を行い、標準画像 Claire に対し良好なワイヤフレームの動きを再現できることを確認した。

第7章 結論

インタラクティブ操作は規格化された処理事項を実行すれば、十分な画質が得られるものでなければならない。本章ではフレームごとにほぼ同じ基準で点の位置を指示しているつもりでも、人間の操作には変動が起こるという問題があることを、実験とデータにより明らかにした。グローバル動きパラメータ検出では剛体仮定を行うが、現実には入力特徴点にノイズが付加されたり、ローカル動きの影響があり、剛体仮定に全面的に依存した3次元処理を行うことは難しい。

特徴点の変動をノイズとみなし2次元画面上での顔形状が時間方向に対して局所的には保存されるという観点から、フィルタの構成を見直し、新たに構造的なフィルタを開発した。従来のように時間方向にLPFを施すだけでは、各点ごとに独立したノイズの成分は除去できるが、時間方向に対する動きの周波数成分も減衰してしまう。これに対し、仮想不動点からの特徴点への距離を拘束することにより、特徴点相互の距離の変動を抑圧することができるようになる。この考えに基づき、顔の構造を時間方向に維持するフィルタを開発した。この構造型フィルタは、特徴点の相互の距離の時間的変動を抑圧しワイヤフレームの形状を保持する効果を持つ。また、本方式では特徴点として、顎以外はローカル動きによる大きい変動の影響を受けない点を用いているため、大きい回転や表情の変化する画像にも適用可能と考えられる。

次に、検出精度に関し、グローバル動きパラメータを求めるために、2つの段階で用いる最小2乗法適用に対して、一般的に使われている3つの仮定に基づき推定値の誤差の分散の評価を行い、Dixonのガイドラインに準拠した形に対し、分散は15%しか増加しないことを明らかにした。

これらのノイズ除去と精度検証により、インタラクティブ操作において、位置として3画素、距離変動では数画素までの許容範囲内で特徴点の入力操作を行うという仕様を設定することができる。これは特徴点位置をマウスで指示入力する操作として、操作量、難易度共に操作者に依存しないシステムが構成できることを示している。

第5章では、インタラクティブ知的符号化におけるインタラクティブ操作が従来、曖昧であったことについての考察に基づき、符号化システムにおけるローカル動き検出以降の部分について、手順の明確化と操作量を評価する検討を行った。口などに対しては、gm→lm方式が、目の開閉に関してはlm→gm方式が有利であることを示し、ローカル動き記述順序の操作回数と難易度の比較によりその裏付けを行った。

次に、ワイヤフレームモデル上の顔部品の頂点に階層的な親子関係を定義し、通常

第7章 結論

の多くの発話シーンで存在する唇の上下動に対し、全点指定の場合に比べ、操作回数と符号量の削減を図る頂点移動方式を開発した。低ビットレート伝送の特徴を生かしたまま、唇両端点に対する0次、0.5次、1次の周期波形をパターン化することにより、閉口やうねるように曲った唇の形状表現におけるインタラクティブ操作量の削減を図った。

開発したインタラクティブ知的符号化の符号化ツールを拡張し、眼鏡処理を用いた符号化を行い、眼鏡フレームに対する肌色による上書き消去処理の操作量の限界が曖昧になることを新たな問題点として指摘した。定量的な判定基準として、テレビゴーストレベルの主観評価実験データを用いて、有限の操作回数で行うための上書き消去処理操作の終了の指標を作成した。具体的には上書き消去処理の不完全性による消え残りが、D/Uで24dB以上になったところで、処理を終了するというインタラクティブ操作に対する基準が作成できた。

以上操作量を考慮して、実用的な時間内に符号化操作を行うことのできるローカル動き検出ツールの開発を行った。本方式は第4章で述べたグローバル動きパラメータ検出ツールと合わせてシステム化し、著者らの他にも技術者が使用し、15種以上の動画系列の符号化が行われている。

第6章では第3、4、5章で提案して来たインタラクティブ知的符号化を、種々の応用を念頭においた実用化実験を行った。

ローカル動きパラメータとグローバル動きパラメータの記述順序の違いによる影響を画質の面から検討した。知的符号化で従来より主に行なわれていた $lm \rightarrow gm$ 記述は対象とする顔動画像のアクションユニットが既知の場合は有効である。現状ではこのアクションユニットが自動でも手動でも完全には検出できないことを考えると、 $lm \rightarrow gm$ 記述は仮想的な顔動画像系列を生成する場合には有効であるが、入力の実画像に忠実な画像合成を行なう場合は、正確な記述ができなくなる。この差異について利害得失の比較検証を行った。

次に、顔動画像の頭部部分の他に肩の動きを別に記述するため、初期画像として、首から下の画像と背景画像との3枚を用いる場合についての符号化実験を行った。CCITT標準画像のClaire約4秒(120フレーム)、Miss America約5秒(150フレーム)の符号化を行い、良好な結果を得た。また、ワイヤフレームモデルと動きパラメータ等に構造化された形で記述された符号化データであるため、復号することなく顔

第7章 結論

のテクスチャを入替えたり、動きを入替えたりすることが出来るが、その検証実験も行った。また、このような標準画像のみでなく、独自に入力した多数の顔動画像に対して符号化を行い、提案して来た符号化システムが汎用的に使用可能であることを検証した。

エージェント研究の画像面の一分野であるヒューマンエージェントにとっては顔動画像の転送、表示が重要な技術となっている。インターネットのホームページのようなハイパーメディアのデータベースにインタラクティブ知的符号化結果を格納し、転送、再生、表示の実験を行った。通常のパーソナルコンピュータでも合成処理に多少時間をかければ、その後はリアルタイムに表示が可能となること、更に OpenGL のグラフィックボードを備えた場合はインターネット回線での遅延を除いて、ほぼ実時間に再生、表示が可能であることが確認された。これらの実験から、通常のパーソナルコンピュータでも実画像に対する実時間の復号、表示が可能で、インタラクティブ知的符号化がヒューマンエージェントを広く実用化していくための基本技術として有効であることが示された。

汎用性に関しては、独自に入力した画像から、15 種以上のものについて符号化を行い、主に顔動画像によるヒューマンエージェントを想定した応用などに使用可能である見通しを得た。また、これらの顔動画像系列の符号化は、筆者以外の技術者が操作して行ったものであり、開発したインタラクティブ知的符号化システムが特殊な処理やチューニングをするようなものでなく、一般性のあるものであることも確かめて来た。

また、再生画像の品質に関しては、従来の低ビットレートの動画像用の波形符号化の代表的存在である H.263 を用い比較実験を行った。画質とビットレートの比較で、提案して来たインタラクティブ知的符号化がほぼ同一かやや優れた性能を示していることが確かめられた。

以上、波形符号化の国際標準化の動きから、次世代知的符号化へ多面的な研究が進展している中で、インタラクティブ知的符号化の提案を行い、実用化を図るための技術的課題をいくつか解決し、符号化実験により、良好な性能の評価と汎用性を確認することが出来た。その中で、波形符号化では、理論的基盤を重視して適応KL変換符号化方式を提案し、良好な符号化性能評価結果を得た。

次に、残された課題について述べる。

第7章 結論

- (1) 適応KL変換符号化方式においては、適応量子化や動画像にたす最適設計手法を確立し、また、画像ブロックに関してはより高次元のブロックでの設計手法を現状あるいは今後数～十数年内に得られるメモリサイズの範囲でいかに行なうかと、という興味深い問題がある。
- (2) インタラクティブ知的符号化に関しては、表示画像サイズを拡大して行く時に処理の増加とともにワイヤフレームモデルの詳細化やインタラクティブ操作画面の拡大化などの検討を行なっていく必要がある。
- (3) 応用面では、人物全体のモデル化、合成等を行なって行く必要があるが、階層的な記述などにより、処理の増加をいかに取り扱っていくかが課題の一つである。

謝辞

本論文をまとめるに当たり、丁寧なるご指導とご教示を賜りました東京工業大学工学部 酒井善則 教授に、謹んで深謝の意を表します。また、ご指導とご討論頂いた東京工業大学 中嶋正之 教授、坂庭好一 教授、植松友彦 助教授、吉田俊之 助教授、電気通信大学 大学院 電気通信学研究科 金子正秀 助教授に謹んで深謝の意を表します。

また、本研究を進めるに際し通信・放送機構、映像分析プロジェクトその他において、長期にわたり、東京大学 工学部 原島博 教授にご指導、ご教示を賜り、深謝致します。また、画像の高効率符号化全般にわたり、全面的にご教示賜った神奈川大学 工学部 斎藤隆弘 教授に心より御礼申し上げます。

本研究の後半部分は主に（株）東芝と通信・放送機構から研究の機会を賜りました。研究にご支援を賜った通信・放送機構の小山森也元理事長、松下操元理事、研究推進部 和田正晴元部長、木原十三男 部長、山田和夫元課長、研究支援を遂行し、最適な環境づくりにご尽力頂いた小林敬幸 元調査役に深く感謝致します。また、通信・放送機構、映像分析プロジェクトのリーダー東京大学 工学部 原島博 教授、サブリーダーの東京工業大学 工学部 酒井善則 教授、神奈川大学 工学部 斎藤隆弘 教授に重ねて感謝致します。また、同サブリーダーの東京大学 工学部 坂内正夫 教授に感謝致します。また、論文の内容に関し、終始多面的な観点から貴重なご教示を賜ったシニア研究フェローの東京大学 工学系大学院 金子正秀 元助教授（元KDD研究所、現電気通信大学）に重ねて心より御礼申し上げます。同じく、シニア研究フェローの成蹊大学 工学部 森島繁生 助教授に感謝致します。また、討論に加わって頂いた、東京工業大学 工学部 吉田俊之 助教授に重ねて感謝致します。また、通信・放送機構、映像分析プロジェクトにおいて、貴重なご討論とご協力頂いた、太田善之研究員、山田悦久研究員、上田剛弘研究員、鄭文涛受入研究員、鄭且根受入研究員の皆様に感謝致します。更に、貴重なご討論とご協力賜った、東京大学 工学部 羽鳥俊光 教授、東京大学 工学部 相澤清晴 助教授に感謝致します。また、同博士課程の永田明德氏（現在、東京工科大学、助手）、同博士課程の苗村健氏（現在、同学部専任講師）、同博士課程の児玉和也氏、同修士課程中川督之氏（現在、NTT）、その他ご協力頂いた学生の方々に感謝致します。同じく成蹊大学 工学部 博士課程坂口竜己氏、同岩沢昭一郎氏、同修士課程 佐藤順氏（現在、（株）東芝）、その他ご協力頂いた学生の方々に感謝致します。

謝辞

本研究の前半部分は主に（株）東芝総合研究所（現在、研究開発センター）、同関西研究所において、研究の機会を賜りました。研究に対し、貴重なご助言、ご指導を賜った神奈川大学 工学部 斎藤隆弘 教授、東京理科大学 理工学部 伊東晋 教授、大阪市立大学 山下一美 教授（現在、関西大学 教授、同計算センター長）、大阪大学 笠原正雄 教授（現在、京都工芸繊維大学、教授）、神戸大学 工学部 田中初一 教授、早稲田大学理工学部富永英義教授に感謝致します。また、研究上の有益なご助言と励ましを頂いた、早稲田大学理工学部数学科 杉山昌平 元教授に感謝致します。

本研究の全ての部分にわたり（株）東芝から研究の機会を賜りました。有益なる各種ご助言とご理解を頂きました、（株）東芝、永井淳 元監査役、下村尚久 専務取締役（現在顧問）、総合研究所 情報システム研究所、久保盛唯 元所長（現在、宇宙開発事業部 顧問）、総合研究所 電子機器研究所 遠藤敬二 元研究主幹、画像機器研究所 高橋貞夫 元所長（現在、芝浦工業大学 教授）、柳町工場 矢幡明樹 技監、情報通信システム事業部 福田武郎 事業部長、研究開発センター 関西研究所 浅川繁元所長（現在、株式会社 関西新技術研究所）、関西研究所 木戸出正継前所長（現在、東芝アメリカ）、情報通信システム事業本部 鈴木秀夫 統括技師長、情報通信システム研究所 南正名 前所長（現在、コンセプトエンジニアリング開発部 部長）、研究開発センター 情報通信システム研究所 杉山文夫 所長、同金子信一元部長、同第6研究所 佐藤幸三前ラボラトリーリーダー（現在 基礎研究所 所長）、同情報通信システム研究所 尾高敏則 ラボラトリーリーダー、同上谷義治研究主務、菊池義浩研究主務、関西研究所 麻田治男所長、に感謝致します。また、有益なるご議論とご助言頂いた、（株）東芝、関西研究所、中村誠ラボラトリーリーダー（現在、神奈川工科大学 教授）、同松井利一元主任研究員、同赤嶺政巳ラボラトリーリーダー、情報通信システム技術研究所 塚原由利子研究員に感謝致します。

更に、お忙しい中、研究の機会を与え、ご支援を頂いた、（株）東芝、小向工場、上村部門長、放送映像機器開発部、萬代慶昭 部長、同佐々木信之主査、同磯部清治 参事、津田幸男参事、東芝A V E 山口晋 技監に感謝致します。

また、研究上の問題点を熱心にご討論頂いた、金沢大学橋本秀雄教授、大阪大学岸野文郎教授、NTTヒューマンインターフェース研究所、大久保栄 研究主幹（現在、通信・放送機構）に感謝致します。

参考文献

- [1]宮川洋, “講座: 帯域圧縮の技術、1. 総論” 電子通信学会誌、vol. 53, No.1, pp.42-43, Jan. 1970.
- [2]宮川洋, “講座: 帯域圧縮の技術、2. 帯域圧縮の基礎理論”, 電子通信学会誌、vol. 53, No.1, pp.44-52, Jan. 1970.
- [3]原島博 監修, “画像情報圧縮”, テレビジョン学会 編, オーム社 1991.
- [4]吹抜敬彦, “画像のデジタル信号処理 増補版”, 日刊工業新聞社、1986.
- [5]樋渡涓二、黒木総一郎, “講座: 帯域圧縮の技術、3. 帯域圧縮の視聴覚心理学”, 電子通信学会誌、vol. 53, No.2, pp.203-212, Jan. 1970.
- [6]W.H. Chen and C.H. Smith, “Adaptive coding of monochrome and color images,” IEEE Trans. Commun., COM-25, 11, pp.1285-1292. Nov. 1977.
- [7]A. Habibi, “Survey of adaptive image coding techniques,” IEEE Trans. Commun., COM-25, 11, pp.1275-1284, Nov., 1977.
- [8]A.K. Jain and P.A. Winz, “Image data compression: a review,” Proc. IEEE 69, 3, pp.349-389, Mar. 1981.
- [9]A. Netravali and J.O. Limb, “Picture coding: a review,” Proc. IEEE 68, 3, pp.366-406, Mar. 1980.
- [10]横沼健雄, “テンソル空間と外積代数”, 岩波講座 基礎数学、線形代数 iv、岩波書店、1977.
- [11]大関和夫, “適応KL変換符号化方式の設計法”, 信学技報、IE92-55, Sept. 1992.
- [12]M. Tasto and P.A. Wintz, “Image coding by adaptive block quantization”, IEEE Trans., Commun., COM-19, No.12, pp.957-971, Dec. 1971.
- [13]M. Tasto and P.A. Wintz, “A bound on the rate distortion function and application to images”, IEEE Trans. Inform. Theory, IT-18, No.1, pp.150-159, Jan. 1972.
- [14]J.J.Y. Huang and P.M. Schulthesis, “Block quantization of correlated Gaussian random variables,” IEEE Trans. Commun. Syst. , CS-11, pp.289-296, Sept. 1963.
- [15]L. Torres-Urgell and R.L. Kirlin, “A new adaptive method for image compression using Karhunen-Löeve transform”, Proc. ICASSP87, 32, 3 pp.1366-1369, 1987.
- [16]M.G. Götze nad Ocylok, “An adaptive interframe transform coding system for images”, Proceedings of ICASSP82, 1, pp.448-451, May, 1982.
- [17]G. Bjøntegaard “Novel method for compressing images using discrete directional transforms”, Proc. SPIE: Visual Communi. And Image Processing '88, 1001, Nov. 1988.

参考文献

- [18]G.W.Wornell and D.H.Staelin, "Transform image coding with a new family of models", Proc. ICASSP88, M2.7, pp.777-780, 1988.
- [19]有本卓, "信号・画像のデジタル処理", 産業図書(1988).
- [20]電子情報通信学会編, "電子情報通信ハンドブック", オーム社(1988).
- [21]佐竹一郎, "行列と行列式", 数学選書 I、p.161, 裳華房(1969).
- [22]L. Torre-Urgell and R.L.Kirlin, "Adaptive image compression using Karhunen-Loeve Transform", Signal Processing, Elsevier, 21, 4, pp.303-313, Dec. 1990.
- [23]H.C. Andrews and C.L. Patterson, "Singular value decomposition (SVD) image coding", IEEE Trans. Commun., Vol. COM-24, pp.425-432, Apr. 1976.
- [24]H.J. Landou and D.Slepian, "Some computer experiments in picture processing for bandwidth reduction", Bell System Technical Journal, 50,5, pp.1525-1540, May-June, 1971.
- [25]大関和夫、上谷義治, "KL変換を用いた適応変換符号化方式", PCSJ87, 4-11, 1987.
- [26]Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design", IEEE Trans. Commun. COM-28, 1, pp.84-95, Jan. 1980.
- [27]手塚慶一、北橋忠宏、小川秀夫, "デジタル画像処理工学", 日刊工業新聞社、1985.
- [28]大関和夫, "適応KL変換符号化方式とその設計法", 電子情報通信学会論文誌、B-I, Vol. J77-B-I, No.2, pp.94-101, Feb. 1994.
- [29]大関和夫、斎藤隆弘、金子正秀、原島博, "マルチメディアメール環境における対話型符号化の基本構成," 1995 信学総大, A-289, 1995.
- [30]川浦康至, "メディアコミュニケーション", 現代のエスプリ、306号, pp.9-19, 1993.1
- [31]H.Li et al. "Image sequence coding at very low bitrates: A review", IEEE Trans. on Image Processing, Vol.3, No.5, pp.589-609, Sep. 1994.
- [32]相澤清晴、原島博、斎藤隆弘, "構造モデルを用いた画像の分析合成符号化方式", 電子情報通信学会論文誌, B-I, Vol. J72-B-I, No.3, pp.200-207, Mar. 1989
- [33]原島博, "次世代画像符号化技術", PCSJ86, パネル1, pp.65-66, 1986.
- [34]原島博, "顔画像の分析合成符号化方式の動向", テレビジョン学会、技術報告、ICS'88-17, pp.19-24, Feb. 1988.
- [35]原島博、相澤清晴、斎藤隆弘, "次世代画像符号化の構想", 電子情報通信学会、技術報告、IE87-1, pp.1-7. April, 1987.
- [36]原島博, "知的画像符号化と知的通信", TV学会誌, Vol.42, No.6, 次世代画像通信小特集、pp.519-525, June 1988.

参考文献

- [37] Masahide Kaneko, "Recent and Current Research on Very Low Bit-Rate Video Coding in Japan," *IEICE Trans. Commun.*, vol.E79-B, no.10, pp.1415-1424, Oct.1996.
- [38]金子正秀, "知的画像情報処理," *テレビ誌*, vol.49, no.6, pp.750-759, 1995.
- [39]金子正秀、羽鳥好律、小池淳, "テキスト情報に対応した口形状変化を有する顔動画像の合成", *電子情報通信学会、技術報告*, IE89-4, pp.29-38.
- [40]金子正秀、羽鳥好律、小池淳, "形状変化の検出と3次元形状モデルに基づく顔動画像の符号化", *電子情報通信学会論文誌 B Vol. J71-B, No.12*, pp.1554-1563 Dec.1988.
- [41]森島繁生, 岡田信一, 原島博, "知的インターフェースのための顔の表情合成法の一検討," *電子情報通信学会論文誌、D-II, Vol. J-73-D-II, No.3*, pp.351-359, Mar. 1990.
- [42] Kazuo Ohzeki, Takahiro Saito, Masahide Kaneko and Hiroshi Harashima, "Interactive Model Based Coding of Facial Image Sequence with a New Motion Detection Algorithm," *IEICE Trans Communi., No.10*, pp.1474-1783, Oct., 1996.
- [43] P.Ekman and W.Friesen, "Facial Action Coding System," Palo Alto, CA: Consulting Psychologists, 1977.
- [44]D. E. Pearson, "Developments in Model-Based Video Coding," *Proceedings of the IEEE*, Vol. 83, No.6, pp.892-906, June 1995.
- [45]H. Li,A.Lundmark and R.Forchheimer, "Image Sequence Coding at Very Low Bitrates: A Review," *IEEE Trans on IP*, vol.3,no.5, pp.589-609, Sept. 1994.
- [46]岡崎透, 崔昌石, 山田寛, 原島博, "顔画像の分析合成システムの実現," *PCSJ90*, 8-8, Oct.,1990.
- [47]U.Cugini, M.Dell'Oca, D.Merelli and P.Mussio , "An Interactive Approach to the Definition of Digital Image Interpretation," *Proc. IEEE Conf. on CVPR* pp168-169, 1983.
- [48]大関和夫, 斎藤隆弘, 金子正秀, 原島博, "インタラクティブ知的符号化システム" *電子情報通信学会論文誌、A, Vol.J80-A, No.8*, pp.1293-1303, Aug., 1997..
- [49]大関和夫、斎藤隆弘、金子正秀、原島博, "インタラクティブ知的符号化におけるグローバル動きパラメータの検出精度," *映像情報メディア学会誌、論文, Vol. 52, No.6*, pp.871-880, 1998.
- [50]H. Li, and R.Forchheimer, "3-D motion estimation in model-based facial image coding," *IEEE Trans. On PAMI*, vol. 15, no.6, pp.545-555, June 1994.

参考文献

- [51] T.S.Huang and A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", Proc. of the IEEE, Vol.82, No.2, pp.252-267, Feb. 1994.
- [52] R.Y.Tsai and T.S.Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," IEEE Trans. On PAMI Vol.6, No1. pp.13-27, Jan. 1984.
- [53] J.Weng, T.S.Huang and N.Ahuja, "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," IEEE Trans. on PAMI Vol.11, No.5., pp451-476, May 1989.
- [54] G.Adiv, "Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects, IEEE Trans. on PAMI, Vol. PAMI-7, No.4, pp384-401, July 1985.
- [55] 広中平祐、他、編集、"数理科学事典", 丸善 1992.
- [56] L.C.W. Dixon, 松原正一訳, "非線形最適化計算法," 培風館, 1974
- [57] M. Covell, M. Withgott, "Spanning the Gap between Motion Estimation and Morphing," Proceedings of 1994 ICASSP, vol 5, pp 213-216, April 19-22 1994.
- [58] 小池淳、金子正秀、羽鳥好律, "顔動画像のモデルベース符号化における頭部の3次元的な動きの推定", 電子情報通信学会論文誌, B-I, Vol.J74-B-I, No.10 pp.789-798, Oct. 1991.
- [59] K.Ohzeki, T.Saito, M.Kaneko, and H.Harashima, "Interactive Model-based Coding for Non-realtime Environment", Proc. of 1996 Picture Coding Symposium, pp.663-668, Mar. 1996.
- [60] 大関和夫, 斎藤隆弘, 金子正秀, 原島博, "インタラクティブ知的符号化における動き解析" 1996 信学総大, A-396, Mar.1996.
- [61] 中川徹、小柳義夫, "最小二乗法による実験データ解析", Up 応用数学選書 7, p42, 東大出版会、1982.
- [62] E.J.Coyle, J-H Lin, and M. Gabbouj, "Optimal Stack Filtering and the Estimation and Structural Approaches to Image Processing", IEEE Trans. on Acoustics Speech and Signal Processing, 37, No.12, pp.2037-2066, Dec.1989.
- [63] Yu-Shan Fong, C.A. Pomalaza-Raez, and X.H. Wang, "Comparison Study of Nonlinear Filters in Image Processing Applications", Optical Engineering 28, No.7, pp749-760, 1989.

参考文献

- [64]Yin, L, Yang, R and Gabbouj, M. : “Weighted Median Filters: A Tutorial”, IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing, 43, No.3, pp.157-192,(Mar.1996).
- [65]H.J.A.M. Heijmans, “Composing Morphological Filters”, IEEE Trans. on Image Processing, 6, No.5, pp.713-723, May, 1997.
- [66]K. Tang, M.Gabbouji, , and J. Astola, “A Novel Nonlinear Vector Filter for Gray-Level Image Filtering”, Proceedings of NSIP95, 25-8, 1995.
- [67]J. Astola, P. Haavisto, and Y. Neuvo, “Vector median filters”, Proceedings of The IEEE, 78, pp.690-710, Apr. 1990.
- [68]L. Alparone, M. Barni, F Bartolini, and V. Cappellini, “Adaptive Weighted Vector-Median Filters for Motion -Fields Smoothing”, Proc. Of ICASSP96, Atlanta, GA. pp2267-2270,
- [69]近藤次郎, “数学モデル－現象の数式化－”, p.93, 丸善, 1976.
- [70]浅野長一郎、江島伸興、李賢平, “基本統計学”, p.258, 森北出版, 1993.
- [71]中村隆英、新家健精、美治泰人、豊田敬, “統計入門”, p.239, 東大出版会, 1984.
- [72]大関和夫、齋藤隆弘、金子正秀、原島 博, “顔動画像のインタラクティブ知的符号化におけるローカル動き検出と眼鏡画像処理”, 情報処理学会、論文誌、Vol. 39 No.7, pp.2134-2148, July 1998.
- [73]大関和夫、齋藤隆弘、金子正秀、原島博, “インタラクティブ知的符号化における動き検出”, 電子情報通信学会、技術報告、MVE95-59, Feb., 1996.
- [74]相澤清晴、 “分析合成符号化におけるモデル構成と表情合成”、電子情報通信学会、技術報告資料、IE87-2, pp.9-15, .
- [75]K.Aizawa and H.Harashima, “Model-based Analysis Synthesis Image Coding (MBASIC) System For A Person’s Face”, Signal Processing: Image Communication Elsevier Science Publishers, pp.139-152, 1989.
- [76]崔昌石, 原島博, 武部幹, “顔の3次元にモデルに基づく表情の記述と合成”, 電子情報通信学会論文誌 A, Vol.J73-A, No.7, pp.1270-1280, Jul. 1990.
- [77]K.Ohzeki, T.Saito, M.Kaneko, and H.Harashima, “Interactive Model-based Coding for Face Metaphor User interface in Network Communications”, Procceding of the International Conference on Intelligent User Interfaces (ACM IUI97), pp.245-248, Jan. 1997.

参考文献

- [78]金子正秀, 小池淳, 羽鳥好律: テキスト情報に対応した口形状変化を有する顔動画像の合成, 電子情報通信学会, 技術報告, IE89-4, Apr. 1989.
- [79]F. Lavagetto, S.Curinga, “Object-oriented Scene Modeling for Interpersonal Video Communication at Very Low Bit-rate”, *Signal Processing: Image Communication* Vol.6, No.5, pp.379-395, 1994.
- [80]荻原和浩, 下等誠巳, “ベジエ曲線によるメガネの形状表現とそのメガネオーダメイドシステムへの応用”, *情報処理学会論文誌*, Vol.36, No.11, pp.2642-2652 Nov. 1995.
- [81]Kazuo. Ohzeki, Masahide. Kaneko, Takahiro. Saito, Hiroshi. Harashima, “Interactive Model-based Coding for Multimedia E-mail Environment”, *Proceedings of the International Workshop on Coding Techniques for Very Low Bitrate Video(VLBV'95)*, A-2, Nov. 1995.
- [82]大関和夫, 斎藤隆弘, 金子正秀, 原島博, “顔動画像のインタラクティブ知的符号化方式, PCSJ95 予稿, 2-3, Oct. 1995.
- [83]松井利一, 平原修三: マルチチャネル構造視覚モデルを用いた複合正弦波知覚特性解析, *電子情報通信学会論文誌 D-II*, Vol.J75-D-II, No.9, pp.1597-1605, Sep. 1992.
- [84]山崎滋, 宮沢寛, “複数ゴースト障害の主観評価定量化における基本評価 DU 比”, *テレビジョン学会誌, 論文*, 第 34, 第 5 号, pp414-419, May, 1980.
- [85]小原正晴, 宮沢寛, 村上純造, 牧野進一, 大関和夫, “アダプター型ゴーストキャンセラーのシステム検討”, *テレビジョン学会, 技術報告 IT52-1*, pp.51-56, Feb. 1982.
- [86]H.Harashima, F.Kishino, “Intelligent Image Coding and Communications with Realistic Sensations-Recent Trends”, *IEICE Trans.*, Vol.E74, no.6, pp. 1582-1592, June 1991.
- [87]Karl Olav Lillevold “Telenor's H.263 Software,”
[http://www.fou.telenor.no/brukere/DVC/h263_software /](http://www.fou.telenor.no/brukere/DVC/h263_software/)

本論文に関する研究業績

(1) 学術論文

- [1]大関和夫、「適応KL変換符号化方式とその設計法」、電子情報通信学会論文誌、B-I, Vol.J77-B-I, No.2, pp94-101, Feb. 1994
- [2]大関和夫、「動画像国際標準符号化(MPEG)」Medical Imaging Technology vol.14, No.3, May pp.237-242, 1996
- [3]Kazuo. Ohzeki, Masahide. Kaneko, Takahiro. Saito, Hiroshi Harashima
"Interactive Model-Based Coding of Facial Image Sequence with a New Motion Detection Algorithm" IEICE Trans. Commun., Vol. E79-B, No.10 pp.1474-1483, Oct. 1996.
- [4]大関和夫、斎藤隆弘、金子正秀、原島 博「顔動画像のインタラクティブ知的符号化システム」電子情報通信学会論文誌、A Vol. J80-A, No.8, pp.1293-1303, Aug. 1997.
- [5]大関和夫、斎藤隆弘、金子正秀、原島 博、「インタラクティブ知的符号化におけるグローバル動きパラメータの検出精度」、映像情報メディア学会誌、論文、Vol.52 No.6, pp.871-880, 1998.
- [6]大関和夫、斎藤隆弘、金子正秀、原島 博「顔動画像のインタラクティブ知的符号化におけるローカル動き検出と眼鏡画像処理」情報処理学会、論文誌、Vol. 39 No.7, pp.2134-2148, July 1998.

(2) 国際会議

- [1]Kazuo Ohzeki, "Three Dimensional Linear Transform Coding", Picture Coding Symposium(PCS77), 2-4, Aug.1977.

本論文に関する研究業績

- [2] Kazuo Ohzeki, "An Adaptive Linear Transform Coding", PCS86, 7.4, 1986.
- [3] Kazuo Ohzeki, "An Adaptive KL Transform Coding Based on a New Transform design Method", PCS88, 1988.
- [4] Kazuo Ohzeki, "Adaptive KL Transform Coding With Motion Compensation For Low Bitrate", MoMuC-1, B.1.5, Dec., 1993
- [5] Kazuo. Ohzeki, Masahide. Kaneko, Takahiro. Saito, Hiroshi. Harashima: "Interactive Model-based Coding for Multimedia E-mail Environment", Proceedings of the International Workshop on Coding Techniques for Very Low Bitrate Video(VLBV'95), A-2 (1995-11).
- [6] Kazuo. Ohzeki, Masahide. Kaneko, Takahiro. Saito, Hiroshi Harashima: "Interactive Model-based Coding for Multimedia E-mail Environment", Proceedings of the International Picture Coding Symposium (PCS'96), Vol.2, pp.663-668 (1996-03).
- [7] Kazuo. Ohzeki, Masahide. Kaneko, Takahiro. Saito, Hiroshi Harashima, "Interactive Model-based Coding for Face Metaphor User interface in Network Communications, International Conference on Intelligent User Interfaces(IUI97), pp.245-248, Jan. 1997.

(3) 専門誌

- [1] Kazuo Ohzeki, "Adaptive Karhunen-Loeve Transform Coding and Its Design Method", Electronics and Communications in Japan Part1:Communications, Vol.78, No.5, pp.10-21, 1995.

本論文に関する研究業績

(4) 口頭発表（国内研究会、全国大会、総合大会等）

- [1]大関和夫、「KL変換を用いた適応変換符号化方式」、Picture Coding Symposium of Japan (PCSJ86), 予稿, 4.7, 1986.
- [2]大関和夫、上谷義治、「KL変換を用いた適応変換符号化方式」、PCSJ87,4-11, 1987.
- [3]大関和夫、上谷義治、「KL変換を用いた適応変換符号化の設計法の比較」
電子情報通信学会、春季全国大会、D-108, 2-100, Mar., 1988.
- [4] 大関和夫、「画像符号化技術の動向」、産学官技術交流懇談会、第61回例会資料、pp.1-9, こうベテクノサロン第12号、神戸商工会議所、Jun. 1992.
- [5]大関和夫、「適応KL変換符号化方式の設計法」、電子情報通信学会、技術報告、IE92-55,pp.45-52, Sep., 1992.
- [6]大関和夫、斎藤隆弘、金子正秀、原島博、「マルチメディアメール環境における対話型符号化の基本構成」、1995 電子情報通信学会、総合全国大会、A-289, 1995.
- [7]大関和夫、斎藤隆弘、金子正秀、原島博、「顔画像のインタラクティブ知的符号化方式」、電子情報通信学会、基礎・境界ソサイエティ大会、A-181, Sep., 1995.
- [8]大関和夫、斎藤隆弘、金子正秀、原島博、「顔動画画像のインタラクティブ知的符号化方式」、PCSJ95 予稿, 2-3, Oct., 1995.
- [8]大関和夫、斎藤隆弘、金子正秀、原島博、「インタラクティブ知的符号化における動き検出」、電子情報通信学会、技術報告、MVE95-59, Feb., 1996.
- [9]大関和夫、斎藤隆弘、金子正秀、原島博、「インタラクティブ知的符号化における動き解析」、電子情報通信学会、総合大会、A-396, Mar., 1996.

本論文に関する研究業績

- [10]大関和夫、斎藤隆弘、金子正秀、原島博、「インタラクティブ知的符号化方式における動き検出の詳細化」、The first Image Media Processing Symposium (IMPS96)予稿、I-6.10, Oct., 1996.
- [11]大関和夫、斎藤隆弘、金子正秀、原島博、「インタラクティブ知的符号化における顔画像の3次元動きパラメータの解析」、電子情報通信学会、技術報告、MVE96-48, Nov., 1996.
- [12]大関和夫、斎藤隆弘、金子正秀、原島博、「インタラクティブ知的符号化における動き情報の入替え」、電子情報通信学会、総合大会、A-16-2, Mar., 1997.

その他の研究業績

(1) 学術論文

- [1]Junzo Murakami, Kazuo Ohzeki and Keisuke Ogi,"A Digitalized Automatic Ghost Cancellor", IEEE Trans. CE-25, pp.555-562,1979
- [2]Masaharu Obara, Takashi Uehara, Shinichi Makino, Tohru Hirata, Kazuo Ohzeki and Junzo Murakami, "A Digital Time Domain Equalizer for Teletext", IEEE Trans. CE-28, No3, pp.447-454, Aug. 1982
- [3]小原正晴、上原隆司、村上純造、大関和夫、「文字多重放送信号用ゴーストキャンセラ」、テレビジョン学会誌、論文、Vol.39., No., 2., pp.170-177, 1985.

国際会議発表リスト

- [1]Kazuo Ohzeki and JunzoMurakami, "Digital Teletext Equalizer Leading To Circuit Integration", IEEE GLOBECOM,48.7, pp1575-1581, Nov., 1984.
- [2]Toshiaki Watanabe, Yuriiko Tsukahara and Kazuo Ohzeki, "Rate-adaptive DCT Coding for Color Pictures", PCS90, 3.13, Mar. 1990.
- [3]Yoshiharu Uetani and Kazuo Ohzeki, "A New Fast Discrete Cosine Transform Algorithm", IEEE 4th DSP Workshop, May, 1990.
- [4]Hideyuki Ueno, Kenshi Dachiku, Kazuo Ohzeki and Fumio Sugiyama, "A Study on Facial Region Detection in the Standard Video Coding Method", 3rd International Workshop on 64kbit/s Coding of Moving Video, 5-2, Sept., 1990.
- [5]Hideyuki Ueno, Kenshi Dachiku and Kazuo Ohzeki ,"Brightness

その他の研究業績

Compensation for Video Phones", PCS91, 13.5, 1991.

[6]Kenshi Dachiku and Kazuo Ohzeki, "Image Coding Using Extra-interpolation VQ", PCS93, 9.2, Mar. 1993.

[7]Kenshi Dachiku, Katsumi Takahashi, Shogo Yamaguchi, Taka-aki Kuratate and Kazuo Ohzeki, "Subband Motion Compensation Extra/Interpolative Prediction Coding at very low Bit Rate", 2'nd International Workshop on Very Low Bitrate Video (VLBV94), 6.2, April, 1994.

専門誌発表リスト

[1]村上純造、大関和夫、牧野進一、桜井優、坂上達郎、大西俊一、小原正晴、
「TVゴーストキャンセラ」、東芝レビュー、36巻7号、pp.625-630, 1981

[2]大関和夫、村上純造、「文字放送用ゴーストキャンセラ」、放送技術、第36巻
5号、pp.451-456, 昭和58年5月

[3]上谷義治、大関和夫、塚本明、「高精度・高速DCT用LSI」、東芝レビ
ューVol.45, No.8, pp.623-626, 1990

口頭発表（国内研究会、全国大会、総合大会等）

[1]大関和夫、「静止画像の帯域圧縮符号化方式（PCM-RL方式）」テレビ
ジョン学会、画像伝送研究会資料 IT-23-2, Oct., 1976.

[2]大関和夫、「静止画像の一符号化方式」電子通信学会、総合全国大会、
1593, 7-177, Mar., 1977.

[3]大関和夫、「三次元線形変換符号化方式」テレビジョン学会技術報告 IT29-

その他の研究業績

7, pp85-90, Nov., 1977.

[4] 大関和夫、「三次元線形変換符号化方式」電子通信学会、総合全国大会、
976, 5-4, Mar. 1978.

[5] 川崎忠通、大関和夫、「デジタル・フィルタ特性近似の一手法」電子通信
学会、通信部門全国大会、246, 1978.

[6] 村上純造、小木恵介、大関和夫、「デジタル化ゴースト消去装置」、電子
通信学会研究会、EMCJ78-37, pp.39-46, Nov.1978.

[7] 村上純造、大関和夫、「ハイブリッド方式ゴーストキャンセラ」テレビジ
ョン学会、技術報告 RE80-6, pp.9-14, Feb., 1980.

[8] 大関和夫、村上純造、大西俊一、小原正晴、「ハイブリッド方式ゴーストキ
ャンセラ」テレビジョン学会全国大会、17-3, pp.393-394, Jul., 1980.

[8] 村上純造、大関和夫、大西俊一、小原正晴、「文字放送用ゴーストキャン
セラ」電子通信学会、総合全国大会、1199, 5-176, Mar., 1981.

[9] 大関和夫、村上純造、大西俊一、小原正晴、「文字放送用ゴーストキャン
セラ」テレビジョン学会全国大会、14-13, 1981.

[10] 大関和夫、村上純造、大西俊一、小原正晴、「文字放送用ゴーストキャン
セラ」テレビジョン学会、技術報告 TEBS73-3, pp.71-75, Aug., 1981.

[11] 小原正晴、宮沢寛、村上純造、牧野進一、大関和夫、「アダプタ形ゴース
トキャンセラーのシステム検討」テレビジョン学会、技術報告、IT52-1,
pp. 51-56, Feb., 1982.

[12] 大関和夫、村上純造、上原隆司、小原正晴、「文字放送用波形等化器の計算
機シミュレーション（基準波形の検討）」テレビジョン学会、技術報告、

その他の研究業績

RE82-24, pp.53-58, Sep., 1982.

- [13]大関和夫、村上純造、「耐ゴーストTV同期分離回路」、電子通信学会、総合全国大会、1189, 5-60, Mar., 1983.
- [14]大関和夫、村上純造、「文字放送用等化器における入力オフセットの影響」、電子通信学会、通信部門全国大会、405, 2-50, 1984.
- [15]関谷邦彦、岡野道商、大関和夫、「カラーディザ画像の符号化方式」、電子通信学会、総合全国大会、1388, 5-216, Mar., 1985.
- [16]上谷義治、大関和夫、「KL変換を用いた適応変換符号化方式」、テレビジョン学会、全国大会、6-6, 1987.
- [17]駄竹健志、大関和夫、「3段階有意画像検出によるフレーム間・フレーム内適応予測符号化」、電子情報通信学会、情報・システム部門全国大会、208, 1-209, 1987.
- [18]上野秀幸、大関和夫、杉山文夫、「動領域の輪郭情報を用いた背景メモリ」、電子情報通信学会、情報・システム部門全国大会、219, 1-220, 1987.
- [19]マイケル クロンク、渡邊敏明、大関和夫、「動き補償におけるブロックマッチングの一検討」、PCSJ87 予稿、6-5, 1987.
- [20]大関和夫、上谷義治、「KL変換を用いた適応変換符号化の設計法の比較」電子情報通信学会、春季全国大会、D-108, 2-100, Mar., 1988.
- [21]渡邊敏明、大関和夫、杉山文夫、「動き補償フレーム間差分DCT符号化方式へのベクトル、スカラー混合量子化の適応」電子情報通信学会、春季全国大会、D-131, 2-123, Mar., 1988.
- [22]駄竹健志、マイケル クロンク、大関和夫、「画素動き補償フレーム間予測

その他の研究業績

符号化」、電子情報通信学会、春季全国大会、D-61, 2-53, Mar., 1988.

[23] 駄竹健志、マイケル クロンク、大関和夫、「画素動き補償符号化方式におけるパラメータの最適化」、テレビジョン学会、全国大会、7-9, 1988.

[24] マイケル クロンク、駄竹健志、大関和夫、杉山文夫、「動画像におけるミスマッチ誤差累積の一検討」、テレビジョン学会、全国大会、7-10, 1988.

[25] 渡邊敏明、駄竹健志、大関和夫、「MC-DCT符号化方式への適応4種差分の適用」、電子情報通信学会、秋季全国大会、D-55, D-1-55, 1988.

[26] 上野秀幸、大関和夫、「背景メモリを用いた動領域抽出」、電子情報通信学会、秋季全国大会、D-69, D-1-69, 1988.

[27] 上野秀幸、大関和夫、杉山文夫、「背景メモリを用いた動領域抽出」、PCSJ88 予稿、7-2, Oct., 1988.

[28] 大関和夫、赤嶺正巳、塚原由利子、「適応KL変換適応符号化方式」、PCSJ88 予稿、7-11, Oct., 1988.

[29] 渡邊敏明、駄竹健志、大関和夫、「適応4種差分-DCT符号化方式」、PCSJ88 予稿、8-2, Oct., 1988.

[30] 駄竹健志、大関和夫、「外挿内挿予測符号化方式」、電子情報通信学会、春季全国大会、D-86, 7-68, Mar., 1989.

[31] 上野秀幸、駄竹健志、大関和夫、「カラー動画TV電話における顔領域検出方式の一検討」電子情報通信学会、春季全国大会、D-92, 7-74, Mar., 1989.

[32] 上谷義治、大関和夫、渡邊敏明、「固体電子スチルカメラ用DCT係数のビット平面符号化方式」、電子情報通信学会、春季全国大会、D-147, 7-129, Mar., 1989.

その他の研究業績

- [33] 駄竹健志、上野秀幸、渡邊敏明、大関和夫、宮崎靖之「ISDN 対応カラー動
画像テレビ電話装置」、電子情報通信学会、春季全国大会、D-233, 7-215, Mar.,
1989.
- [34] 赤嶺正巳、塚原由利子、大関和夫、「適応ビット配分を有する適応KL変換
適応符号化方式」、電子情報通信学会、秋季全国大会、D-42, 6-42, 1989.
- [35] 渡邊敏明、大関和夫、「平均値を用いた交流成分予測方式の一検討」、PCSJ89
予稿、2-2, Oct., 1989.
- [36] 大関和夫、赤嶺正巳、塚原由利子、「エントロピー符号を用いた適応KL変
換適応符号化方式」、PCSJ89 予稿、4-1, Oct., 1989.
- [37] 上野秀幸、大関和夫、「顔領域検出を用いた動画像符号化方式」、PCSJ89
予稿、7-15, Oct., 1989.
- [38] 上谷義治、大関和夫、「DCT 新高速アルゴリズム」、電子情報通信学会、
春季全国大会、A-191, 1-191, Mar., 1990.
- [39] 渡邊敏明、大関和夫、「レート適応型DCT符号化方式の性能改善」テレビ
ジョン学会、年次大会、17-6, 1990.
- [40] 上谷義治、大関和夫、「DCT 新高速演算アルゴリズムの性能」、電子情報
通信学会、DSP 研究会、技術報告、DSP90-74, pp.67-72, Aug., 1990.
- [41] 塚原由利子、赤嶺正巳、大関和夫、「適応KL変換符号化方式における変換
選択方法の検討」、電子情報通信学会、秋季全国大会、D-188, 6-190, 1990.
- [42] 上野秀幸、駄竹健志、大関和夫、「人物領域検出を用いた輝度補正の検討」
情報通信学会、秋季全国大会、D-230, 6-232, 1990.

その他の研究業績

- [43]大関和夫、塚原由利子、赤嶺正巳、「エントロピーを考慮した変換行列設計を行う適応KL変換符号化方式」PCSJ90 予稿、4-10, Oct., 1990.
- [44]上野秀幸、駄竹健志、大関和夫、「顔領域検出符号化におけるビット配分方式の比較検討」PCSJ90 予稿、8-13, Oct., 1990.
- [45]上谷義治、大関和夫、塚本明、「高精度・高速DCT用LSI」、画像電子学会、年次大会、37, 1991.

特許リスト(登録済みのもの)

番号	特許登録番号 登録日	発明名称	発明者
1	特許第1145273号 昭和58年5月12日登録	画像伝送方式	大関和夫
2	特許第1165161号、 昭和58年8月26日登録	カラー画像伝送方式	大関和夫、小木恵介
3	特許第1187255号 昭和59年1月20日登録	符号化方式、	大関和夫、
4	特許第1217120号 昭和59年7月17日登録	画像信号伝送方式、	大関和夫、小木恵介、竹内敬二、 矢幡明樹
5	特許第1268201号 昭和60年6月10日登録	垂直同期信号分離回路、	大関和夫、大西俊一、小原正晴、
6	特許第1278125号 昭和60年8月29日登録	画像信号伝送方式、	大関和夫、小木恵介、竹内敬二、 矢幡明樹
7	特許第1339231号 昭和61年9月29日登録	垂直同期信号分離回路、	大関和夫、大西俊一、小原正晴、
8	特許第1425647号 昭和63年2月15日登録	同期検波回路、	大関和夫、村上純造、大西俊一、
9	特許第1443209号 昭和63年6月8日登録	画像信号処理方式、	大関和夫、
10	EPC0156480 April 20, 1988	Timing control signal generator for teletext signal waveform	K. Ohzeki,
11	特許第1443266号、 昭和63年6月8日登録	、伝送路歪検出装置、	大関和夫、村上純造、大西俊一、 小原正晴
12	特許第1451417号 昭和63年7月25日登録	信号処理装置、	大関和夫、大西俊一、小原正晴、
13	特許第1451417号 昭和63年7月25日登録	信号処理装置、	大関和夫、大西俊一、小原正晴、
14	特許第1482367号 平成1年2月27日登録	自動等化器、	大関和夫、大西俊一、
15	USP4833535 May 23, 1989	Image Transmission Apparatus	K.Ohzeki, F.Sugiyama, K.Dachiku, H.Fujiwara, T.Watanabe,

番号	特許登録番号 登録日	発明名称	発明者
16	特許第1537148号 平成1年12月21日登録	テレビジョンゴースト 消去装置、	大関和夫、村上純造、大西俊一、 小原正晴、
17	特許第1575432号、 平成2年8月24日登録	テレビジョンゴースト 消去装置、	大関和夫、村上純造、大西俊一、 小原正晴
18	特許第1577862号 平成2年9月13日登録	波形等化器、	大関和夫、村上純造、大西俊一、 小原正晴、
19	特許第1672191号 平成4年6月12日登録	垂直同期信号始点検出 回路、	大関和夫、大西俊一、小原正晴、
20	特許第1672192号、 平成4年6月12日登録	垂直同期信号始点検出 回路	大関和夫、大西俊一、小原正晴、
21	特許第1690392号 平成4年8月27日登録	位相同期回路、	大関和夫、小原正晴、
22	特許第1727636号 平成5年1月19日登録	波形等化器、	大関和夫、上原隆司、小原正晴、
23	特許第1814571号 平成6年1月13日登録	自動波形等化器、	大関和夫、
24	特許第1838043号 平成6年4月11日登録	文字放送信号受信装 置、	大関和夫、
25	特許第1874431号 平成6年9月26日登録	多階調画像信号の符号 化装置、	大関和夫、関谷邦彦、岡野道商
26	特許第2603931号 平成9年1月29日登録	画像伝送装置	大関和夫、駄竹健志、杉山文夫
27	特許第2637093号 平成9年4月25日登録	画像符号化方式、	大関和夫
28			
29			
30			
31			