

最近の画像圧縮技術：視覚研究に期待するもの

緒形昌美

ソニー（株）メディアプロセッシング研究所

〒141-0001 東京都品川区北品川 6-7-35

1. はじめに

視覚情報は、動物が生活していく上での状況判断に必要なものであるが、我々人間にとってはさらに、鑑賞の対象として“心を潤す”という大きな役割を担っている。画像・映像が我々を魅了するという事実は、今なおテレビ、映画などが大きな市場を提供していることから明らかである。画像圧縮技術は、このような画像情報の利用範囲・利用形態を大きく広げ、新たなマーケットを開拓する可能性を持つものとして多くの企業が盛んに研究開発を行っており、これらの成果はビデオ CD (Compact Disk) や DVD (Digital Versatile Disk) などの実用化に結び付いている。

ところで、MPEG (Moving Picture Expert Group) などの国際標準化作業も含めた集中的な技術開発を通して近年明らかになってきたことは、圧縮効率と汎用性、および実装の容易さなどのバランスを考慮した場合、少なくとも動画像の圧縮方式としては DCT (Discrete Cosine Transform) とブロックマッチングによる動き補償 (以下 BKMC) の組み合わせがほぼ最適な方式であり、またその性能も飽和状態に近づきつつあるということである。しかし、この技術的な成熟度が必ずしもユーザーの満足度に対応しているわけではなく、特に DCT+BKMC 方式特有のブロックノイズは、

人間の感覚にとっては不自然な劣化であるとして、その改善、あるいはさらなる圧縮効率の向上を望む声は依然強い。また最近では、汎用性よりも機能性を重視した圧縮方式に対する関心が高まってきており、ビットストリーム上における編集作業や画像検索など、所望する機能の実現のしやすさといった観点からの検討が行われている。

本稿では、上記のようなさまざまな目的のために現在研究開発が進められている代表的な画像圧縮技術について概説するとともに、今後、圧縮効率の向上や要求される機能性を実現していく上で、視覚研究に期待するものについて、著者なりの意見を述べる。

2. DCT+BKMC 方式

2.1 基本技術

本節では、動画像圧縮方式として現在最も広く用いられている DCT+BKMC 方式の概要を説明する。詳細については文献¹⁾などを参照されたい。また、静止画像の圧縮方式である JPEG²⁾などは、以下に述べる動画像圧縮方式からフレーム間処理、すなわち動き補償処理を除いたものと考えられる。

DCT+BKMC 方式の符号化器の基本的構成を図 1 に示す。破線内部はローカルデコーダと呼ばれ、復号器側と同じ動き補償処理を行うために設けられている。実際の復号器の構

成は、基本的にはこのローカルデコーダにランレングス符号化、VLCの逆処理を行う部分が付加されたものとなっている。以下に、各処理について説明する。

2.1.1 DCT (Discrete Cosine Transform)

DCTはフーリエ変換などと同様、画像データを周波数成分に分解する正規直交変換であり、次式によって定義される。

$$f(u,v) = \frac{1}{4} C(u)C(v) \times \sum_{i=0}^7 \sum_{j=0}^7 x(i,j) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16}$$

$$C(u)C(v) = \begin{cases} \frac{1}{\sqrt{2}} & u,v=0 \\ 1 & u \neq 0, v \neq 0 \end{cases} \quad (1)$$

通常の画像圧縮では、式(1)に示されているように、入力された画像を8画素×8画素のブロックに分割し、各ブロック毎にDCTを施し

て64個の変換係数を得る。もともと画像データは空間的相関が高いため、DCTによって得られる変換係数は低周波数に対応するものほど大きな値をとることになるが、これによってデータの分布の偏りが大きくなり、冗長度が削減されて平均的な情報量(エントロピー)が減少する。冗長度の削減能力のみを考えれば、理論的にはK-L(Karhunen-Loeve)変換が最も優れているが、処理時間の問題があり、DCTは実用レベルで最も冗長度削減能力が高いものとして広く用いられるようになっている。

2.1.2 動き補償 (MC: Motion Compensation)

時系列的に入力される各画像は、その順序にしたがって図2に示すようにIピクチャ、Pピクチャ、Bピクチャのいずれかのピクチャタイプが割り当てられる。IピクチャではDCTによる空間的冗長度の削減のみが行われるのに対して、Pピクチャ、およびBピク

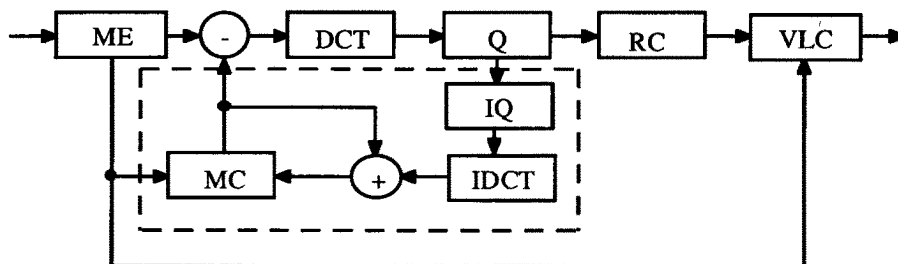


図1 DCT+BKMC方式の基本構成。ME: 動き検出, MC: 動き補償, Q: 量子化処理, IQ: 逆量子化処理, IDCT: 逆DCT, RC: ランレングス符号化

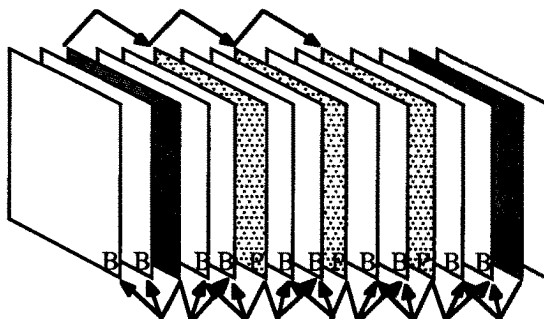


図2 ピクチャ符号化タイプと予測方向

チャでは、さらに時間的冗長度を削減するために、DCTを行う前に次式に示すような動き補償が施される。

$$y_{ijk} = x_{ijk} - x_{i+m, j+n, l} \quad (2)$$

ここで、 x_{ijk} は時刻 k のフレームにおける位置 (i,j) の画素の値であり、 $x_{i+m, j+n, l}$ は時刻 l のフレームにおいて x_{ijk} に対応する画素の値である。すなわち (m,n) が動きベクトルに相当する。 y_{ijk} は予測誤差であり、P ピクチャ、B ピクチャでは、この予測誤差に DCT が施される。

動きベクトルは通常、16 画素×16 画素のマクロブロックと呼ばれるブロック単位で求められる。図2に示すように、P ピクチャーでは過去のフレームからの予測が行われ、B ピクチャーでは過去のフレーム、未来のフレーム、あるいはその両方からの予測の中で最も予測誤差が小さくなるものが選ばれる。このとき、参照するフレーム（過去のフレーム、あるいは未来のフレーム）は線形補間によって画素間の値が補間され、半画素精度の動きベクトルが求められる。

2.1.3 量子化、逆量子化

DCTによって得られた DCT 係数 $f(u,v)$ は、次式に示すような量子化処理が施される。

$$f'(u,v) = \frac{f(u,v)}{Q} \quad (3)$$

ここで、 Q は量子化スケールであり、この値が大きいく程再構成された画像の劣化が大きくなる。逆量子化処理では、量子化係数 $f'(u,v)$ に量子化スケール Q を積算することにより DCT 係数を復元する。

なお、量子化スケールはすべての DCT 係数に対して同じものを用いる必要はなく、視覚系の周波数特性を考慮して重みづけがなされる場合もある。

2.1.4 ランレングス符号化/VLC(Variable Length Coding)

量子化処理の施された量子化係数は、各 DCT ブロック毎に図3に示すような順序で走査され、ランレングス符号化、および2次元 VLC が施される。

ランレングス符号化は量子化後の係数の多くがゼロの値となることを利用したものであり、図4に示すように、走査順に現われる係数列を、非ゼロの係数までの連続するゼロの数（ゼロラン）とその非ゼロの係数値の組として表わすものである。また、係数列における最後の非ゼロ係数の直後では EOB (End of Block) という特別のシンボルを符号化する。これによって、以後のゼロ係数を符号化する必要がなくなり、圧縮効率が向上する。

ランレングス符号化の施されたゼロランと非ゼロ係数値の組は、VLC テーブルと呼ばれるテーブルによって、ビット列に変換される。VLC は、出現頻度の高いシンボルに対しては短いビット列、出現頻度の低いシンボルに対しては長いビット列を割り当てるもので

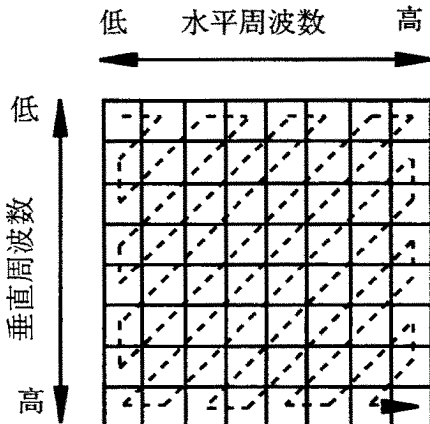


図3 DCT係数の走査順

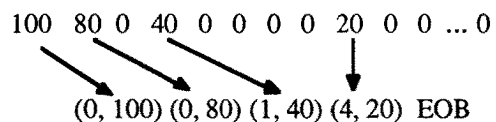


図4 ランレングス符号化

あり、ハフマン符号と呼ばれる符号が多く用いられている。同様に、動き検出によって求められた動きベクトルもVLCによって符号化される。

2.2 MPEG-4

MPEG-4⁶⁾は、現在標準化作業が進められているオブジェクトベースの圧縮方式であり、任意の形状を持つ領域の時系列（オブジェクト）を処理することが可能になっている。たとえば、図5に示すような背景とその前に位置する人物からなる画像シーケンスを圧縮する場合、従来の方式では、画像の内容にかかわらず長方形の2次元配列を処理対象としていたのに対し、MPEG-4では、人物の領域と背景領域を独立に圧縮することができる。このため、符号化されたビットストリーム上では各オブジェクトを独立に扱うことができ、さまざまなアプリケーションへの応用が期待できる。各領域の圧縮を行う基本的技術はやはりDCT+BKMCであるが、任意形状を扱うために、領域形状の圧縮処理、任意形状領域の内部情報の圧縮処理などの新しい技術が導入されている⁹⁾。

なお、このような任意形状をしたオブジェクトデータをどのように用意するかはMPEG-4の規格範囲外であり、オフラインでの処理

が許されるのであればクロマキー等の方法を用いることもできる。ただし、MPEG-4の応用範囲をより広げるためには、自動的なセグメンテーション手法の実現が望まれる。セグメンテーションは画像認識などの基本的な前処理として古くから研究が行われてきているが⁷⁾、上記のような理由により、主に動画を対象に、近年画像符号化分野の研究者の間でも研究が始められている⁸⁻¹⁰⁾。

2.3 DCT+BKMCの問題点

DCT+BKMC方式では処理がブロック単位で行われるため、量子化誤差が隣接ブロック間で無相関となる。特に、低周波数成分の量子化誤差が大きい場合には、復元された画像上にブロック状のパターン（ブロックノイズ）が現われる。ブロックノイズを軽減させるための手段としては、動き補償を行う際に、隣接ブロックの動きベクトルを用い、各動きベクトルによって得られる参照ブロックの重み付平均として予測を行うオーバーラップ動き補償⁹⁾などがあるが、必ずしも十分な効果が得られていない。

また、高周波数係数の量子化による劣化としてはモスキートノイズと呼ばれるものがあり、エッジ等の広い帯域を有する信号の近傍に蚊がまわりつくように発生する。

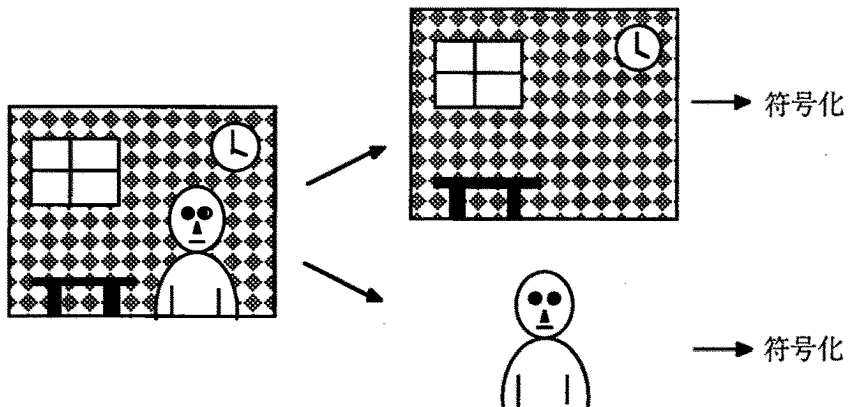


図5 オブジェクトベース符号化方式の概念

3. その他の圧縮技術

本章では、DCT+BKMC 以外の画像圧縮技術について、1) 空間的冗長度を削除するための技術、2) 時間的冗長度を削除するための技術に分けて紹介する。

3.1 空間情報の圧縮技術

3.1.1 ウェーブレット符号化法

ウェーブレット変換は、DCTと同様、空間的冗長度を削減するために用いられる。その基本的操作は、ローパスフィルタとハイパスフィルタによる畳み込み積分とダウンサンプリングの繰り返しであり、これを用いたウェーブレット符号化法^{11,12)}は、原理的にブロックノイズを発生しない。実際、静止画像に対してはDCTよりも高い画質が得られることが報告されており、MPEG-4の暫定モデルにも texture coding modeとして採用されている。また、ウェーブレット変換後のデータ構造は、画像の多重解像度表現になっており、空間スケラビリティを実現する符号化方式としても期待されている。空間スケラビリティとは、1つのビットストリームの中に空間解像度の異なる複数の画像情報を効率よく埋め込む機能であり、ビットストリームの1部を復号することで、演算量に見合った解像度の画像を再構成することができる。これにより、伝送容量の異なる様々なネットワークが接続された環境や、性能の異なる複数の復号器を対象とする場合などでも1つのビットストリームで対応することができるようになる。

ただし、ブロックマッチングによる動き補償との相性が悪く、Pピクチャ、Bピクチャに対する圧縮性能はDCTの方が優れている。オーバーラップ動き補償との組み合わせによってその性能を大きく改善できることも報告されているが¹³⁾、DCT+BKMC方式を凌駕するには至っていない。

3.1.2 フラクタル符号化法

フラクタル符号化法¹⁴⁾は、画像の持つ自己相似的な構造に着目したもので、符号化処理では、入力された画像を近似するための縮小

変換が次のように求められる。すなわち、画像をレンジブロックと呼ばれる重複しないR画素×R画素のブロックに分割し、各レンジブロックごとに、対応するD画素×D画素のブロック（ドメインブロック）を同じ画像内から検出する。このドメインブロックは、縮小変換を施すことによって、現在注目しているレンジブロックを最もよく近似できるものが選択される。各レンジブロックは、得られた縮小変換パラメータによって表現される。

復号器では、任意の初期画像に対して、伝送されてきた変換パラメータによる縮小変換を繰り返し適用する。この反復処理が、原画像のよい近似に収束していくことは、不動点定理、コラージュの定理によって保証されている。縮小変換としては、アフィン変換が多く用いられている。

フラクタル符号化法は、符号化処理における縮小変換の生成に多くの演算を必要とするが、復号処理は高速に実現できる。また、静止画像に対しては、高い圧縮率において、DCTを用いたJPEG方式よりも良い画質が得られることが報告されている。さらに、復号時にレンジブロックの大きさを変えることによって、同じビットストリームから拡大・縮小などを施した再生画像を生成することができる¹⁵⁾。

3.1.3 Matching pursuit

Matching pursuitは、動き補償によって得られる予測誤差の画像（差分画像）を圧縮するための技術であり、特に低レートではDCTを上回る性能を示している。この手法では、あらかじめ小さなパターン（アトム）を多数用意しておき、入力画像の局所領域ごとに、最もよく対応するアトムを検出する。アトムの検出は、入力された画像において、局所的なエネルギーの大きい部分から順に行っていく。また、対応の度合は小領域とアトムの内積によって計算される。注目小領域の中心座標、選択されたアトムのインデックス、および内積値を符号化することによって圧縮を行

う。Neffらは、アトムとして周波数、位相、広がり異なる20個の1次元ガボア関数(実関数)を用い、これらを組み合わせて生成される400個の2次元パターンをアトムとして用いている¹⁶⁾。これは400個の基底関数を用いて画像を展開していることに相当するが、その数はDCT(64個)や、ウェーブレット変換(N回分割で $3N+1$ 個)と比較しても非常に多く、差分画像上に現われる非定常なパターンを効率的に表現することが可能となる。

なお、matching pursuitでは、最適なアトムを検出するために多くの演算を必要とすることが1つの問題点となっている。また、DCTやウェーブレットと異なり、各アトムの位置情報を明示的に伝送する必要性があり、伝送レートが高くなった場合に、この情報量が圧縮効率を下げる原因になることも考えられる。

3.2 時間情報の圧縮技術

3.2.1 ワーピング

現在、最も広く用いられている動き補償方式では、動きを2次元平行移動のみによってモデル化しているため、画像上に、回転、拡大(進出)、縮小(後退)などの動きを示す物体が投影されている場合には、当然予測誤差は大きくなり圧縮効率は低下する。ワーピング^{17,18)}はこのようなさまざまな動きにも対応するための技術であり、ブロックノイズの低減効果も併せ持つ。この手法では、画像を適当な形状の小片(パッチ)に分割し、各パッチの頂点位置における2次元動きベクトルを算出し、これらの動きベクトルからパッチ内の各画素に対応する動きベクトルを内挿して求める。内挿の方法としては、アフィン変換や双一次変換などが用いられる。これにより、動き補償の性能が向上することは明白であるが、動きを表わすパラメータが増加してしまい全体的な圧縮効率が低下してしまう可能性もある。また、演算量の増加も指摘されており、その高速化も検討されている¹⁹⁾。

3.2.2 GMC(Global Motion Compensation)

GMC^{20,21)}は、カメラワーク等によって生じる画像全体の動きを補償するための手法であり、技術的には画像全体を1つのパッチとしたワーピングと考えられる。これによる付加的パラメータは1フレームに1組あればよいので、カメラワークの多い映像素材等が圧縮対象である場合には大きな効果が期待できる。また、全体的な動きを補正した後に局所的な動きを補正するといった戦略は、理論的には妥当なものと考えられる。しかし全体として、それを実装するためのコストに見合うだけの効果が得られるか否かは定かではなく、その有用性は対象とする画像、あるいはアプリケーションに強く依存するものと考えられる。

4. 視覚研究への期待

本章では、さらなる圧縮効率の向上や新たな機能性の実現に向け、著者が視覚研究に期待するいくつかの項目について述べる。

4.1 視覚系における符号化メカニズム

DCT+BKMC方式の性能の高さは、DCT、およびBKMCそれぞれの空間的冗長度、時間的冗長度の削減能力に加え、それらの組み合わせの相性のよさにあると思われる。従って、2.3節で述べたようなDCT+BKMCの抱える問題を克服し、飛躍的な圧縮効率の向上を実現するためには、全体的な構成から見直すことも必要になると思われる。このとき、コンピュータビジョンなど視覚情報を扱う他の分野と同様、視覚系のメカニズムを模倣することが、1つの有効なアプローチとして期待される。例えば、視覚系における空間的冗長度を削減する手段としては、ブロックベースのDCTよりも畳み込み演算に基づいたウェーブレット変換の方が妥当であるように思われるが、このような空間的変換に対して、どのような時間軸処理が組み合わされているのかが興味深いところである。

4.2 画質の客観的評価尺度

主観的評価と高い相関を持つ客観評価尺度の必要性は従来から指摘されており、いくつかの試みもなされているが^{22,23)}、いまのところ広く認知される方法は確立されておらず、MPEGなどの標準化活動においても、原画像と再構成画像の画素単位の差分を表わすSNR (Signal to Noise Ratio)が用いられているのが現状である。このSNRは単純であり、また類似した方式どうしを比較する場合にはかなり主観評価を反映した結果が得られるが、DCTによるブロックノイズのような構造的な劣化を評価することができない。さらにDCTとウェーブレット変換のように、異なる種類の劣化を生じるような手法を比較することも難しい。

また、信頼性の高い客観評価尺度が利用できれば、符号化処理における画質制御に適用することも可能である。現在、パッケージメディアなど、オフラインで符号化処理を行う場合、与えられた画質の最適化を行うために、多くの時間をかけて手動で画質制御（量子化ステップの調整）を行っているが、客観的評価尺度を組み込むことにより、その自動化、あるいは省力化を実現できる可能性がある。

4.3 画像解析

画像解析は、これまでコンピュータビジョンなどの分野で研究が進められてきたが、近年、画像情報をより扱い易いものにするを目的に、画像符号化分野からのアプローチが始められている。たとえば、2.2節で紹介したMPEG-4への応用を目的としたセグメンテーションの自動化、効率的な画像検索を実現するための特徴量抽出などはその代表的なものと言える。このことは、これらの技術に対する需要が具体的な形で高まってきていることを意味しており、その研究開発においては、これまで同様視覚研究の貢献が期待される。

5. おわりに

本稿では、最近の画像圧縮技術の概要を解説し、今後の画像符号化分野において視覚研究に期待するものについて述べた。

量子化処理を含む非可逆的な画像圧縮技術の性能は、最終的には我々人間の感覚よって判断されるべきものであり、その開発に際しては常に視覚系の特性を考慮していく必要があることは言うまでもないが、さらに圧縮方式全体のバランスといった視点から、視覚メカニズムの解明が圧縮効率の飛躍的向上に寄与することが期待される。また、今後人間と画像情報の関わり方が多様化するにつれ、画像解析処理などの必要性が高まることが予想され、この面でも視覚研究における成果が注目される。

文 献

- 1) B. G. Haskell, A. Puri and A. N. Netravali: Digital video: Introduction to MPEG-2. Chapman and Hall, New York, 1997.
- 2) テレビジョン学会 (編) : MPEG. オーム社, 1996.
- 3) W. B. Pennebaker and J. L. Mitchell: JPEG. Van Nostrand Reinhold, New York, 1993.
- 4) L. Chiariglione: MPEG and multimedia communications. *IEEE Transactions on Circuits and Systems for Video Technology*, 7, 5-18, 1997.
- 5) T. Sikora: The MPEG-4 video standard verification model. *IEEE Transactions on Circuits and Systems on Video Technology*, 7, 19-31, 1997.
- 6) MPEG Video Group: MPEG-4 video verification model version 8.0. ISO/IEC JTC1/SC29/WG11 N1796, 1997, draft in progress.
- 7) N. R. Pal and S. K. Pal: A review on image segmentation techniques. *Pattern Recognition*, 26, 1277-1294, 1993.
- 8) P. J. Mulroy: Video content extraction: Review of current automatic segmentation algorithms. *Proceeding of Workshop on Image Analysis for Multimedia Interactive Services*, 45-50, 1997.
- 9) 苗村昌秀: スポーツ番組における事前知識情報を用いたモルフォロジカルセグメンテーション. 映像メディア処理シンポジウム IMPS97予稿集, 47-48, 1997.

- 10) F. Ziliani and F. Moscheni: Kalman filtering motion prediction for recursive spatio-temporal segmentation and object tracking. *Proceeding of Workshop on Image Analysis for Multimedia Interactive Services*, 63-68, 1997.
- 11) S. A. Martucci, I. Sodagar and T. Chiang: A zerotree wavelet video coder. *IEEE Transactions on Circuits and Systems for Video Technology*, 7, 109-118, 1997.
- 12) J. M. Shapiro: Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41, 3445-3462, 1993.
- 13) M. Ohta and S. Nogaki: Hybrid picture coding with wavelet transform and overlapped motion-compensated interframe prediction coding. *IEEE Transactions on Signal Processing*, 41, 3416-3424, 1993.
- 14) M. F. Barnsley and L. P. Hurd: *Fractal Image Compression*. A K Peters Ltd., 1992. 蔡 東生, 江守正多 (訳) : マルチメディアフラクタル画像圧縮. トップラン, 1995.
- 15) 勝山 栄, 木本伊彦, 谷本正幸: フラクタル符号化を利用する解像度変換. 画像符号化シンポジウム PCSJ94予稿集, 89-90, 1994.
- 16) R. Neff and A. Zakhor: Very low bit-rate video coding based on matching pursuits. *IEEE Transactions on Circuits and Systems for Video Technology*, 7, 158-171, 1997.
- 17) 中屋雄一郎, 中井敏久: モバイル映像メディアにおける画像符号化の動向. 映像情報メディア学会誌, 51, 1612-1619, 1997.
- 18) V. Seferidis and M. Ghanbari: General approach to block-matching motion estimation, *Optical Engineering*, 32, 1464-1474, 1993.
- 19) 中屋雄一郎, 鈴木芳典: ソフトウェアコーデックのための高速ワーピング処理. 画像符号化シンポジウム PCSJ96予稿集, 55-56, 1996.
- 20) 上倉一人, 如沢裕尚, 渡辺 裕, 小寺 博: 動画画像符号化におけるグローバル輝度補償法について. 画像符号化シンポジウム PCSJ96予稿集, 31-32, 1996.
- 21) H. Jozawa, K. Kamikura, A. Sagata, H. Kotera and H. Watanabe: Two-stage motion compensation using adaptive global MC and local affine MC. *IEEE Transactions on Circuits and Systems for Video Technology*, 7, 75-85, 1997.
- 22) 宮原 誠, 小谷一孔, 堀田裕弘, 藤本 強: 客観的画質評価尺度 local feature の考慮と汎用性. 電子情報通信学会論文誌 (B-I), J73-B-1, 208-218, 1990.
- 23) 櫻井篤志, 桐澤 深: 符号化画像の画質評価. 画像電子学会誌, 21, 520-510, 1992.