社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

画素を単位とした動き補償に関する検討

— 水平方向に限定した予備検討 —

内田　誠一†　　迫江　博昭†

† 九州大学大学院システム情報科学研究院
E-mail: †{uchida,sakoe}@is.kyushu-u.ac.jp

あらまし　動画像圧縮においては，圧縮率向上のために動き補償が利用される．ブロックやパッチといった小領域を単位とする従来の動き補償法では，動き情報に要する符号量が少くて済む反面，小領域内の細かい動きには追従できず残差の極小化能力には限界があった．そこで本研究では，画素を単位とした動き補償について検討する．動き情報符号量の増加を如何に抑えるかが課題となるが，隣接画素間の動きを相互に制約することで解決を図る．本報告では，第一段階として，水平方向に限定した場合について考察を行う．実験により，動きの大きな動画像の圧縮における有効性を示した．
キーワード　動き補償, 画素対応, 動画像圧縮, 最適化

A preliminary study of pixel-based motion compensation

Seiichi UCHIDA† and Hiroaki SAKOE†

† Faculty of Information Science and Electrical Engineering, Kyushu University
E-mail: †{uchida,sakoe}@is.kyushu-u.ac.jp

Abstract　For efficient video compression, a pixel-based motion compensation technique where motion is determined at every pixel is investigated. The pixel-based motion compensation technique is expected to provide better performance on the minimization of prediction error than conventional block-based motion compensation techniques. In order to suppress the increase of motion information, several constraints are imposed on the motions of neighboring pixels. Experimental results have shown that the present technique can attain higher compression rates than a block-based technique for image sequences with large motion.
Key words　motion compensation, pixel correspondence, video compression, optimization

1. ま え が き

動画像圧縮においては，圧縮率の向上を目的として一般に動き補償が用いられる [1], [2]．この動き補償とは，連続する2フレーム間の動きを同定する技術である．その動きに基づいてフレーム間予測を行えば，得られる予測残差は0付近に集中したものとなる．従って，その後の適当なエントロピー符号化により，高い圧縮率を得ることができる．

動き補償を用いた場合，予測残差情報に加えて動き情報を伝送しなければならない．よって全体として圧縮効率を高めるためには，予測残差の符号量と動き情報の符号量を共に小さくする必要がある．この要求に対し，ブロックやパッチなどの小領域に一様な(すなわち平均的な)動きを求める方法がある．これは動き情報符号量の削減を狙ったもので，MPEGなどの標準符号化方式で広く採用されている．しかし，こうした小領域ベースの動き補償法では小領域内に複数の動きや非剛体変形が発生すると残差が大きくなり，予測残差の符号量が大きくなってしまうことがある．

本報告では，画素を単位とした動き補償手法を提案する．画素を単位とすることで，動きに対する柔軟性が増し，小領域ベースの手法では問題となるような複雑な動きが発生した場合についても，予測残差符号量を小さくできる．一方で，画素単位化に伴う動き情報の符号量の増加を如何に抑制するか課題となる．これについては，隣接画素の動きを相互に制約し，その差分値のダイナミックレンジを極力小さくすることで対処する．

動画像中の物体の動きは2次元的なものであるが，本報告では画素単位の動き補償法に関する検討の第一段階として，水平方向(左右方向)の動き補償に限定した場合について考察を行う．この場合の動き補償は，連続する2フレームの同一ライン間の画素対応決定問題として定式化される．動きの2次元的拡張に

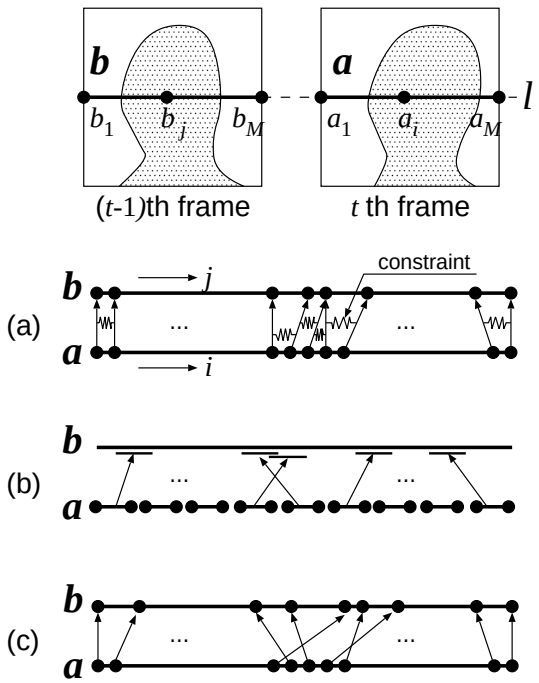

図1　同一ライン間マッチングによるフレーム間予測

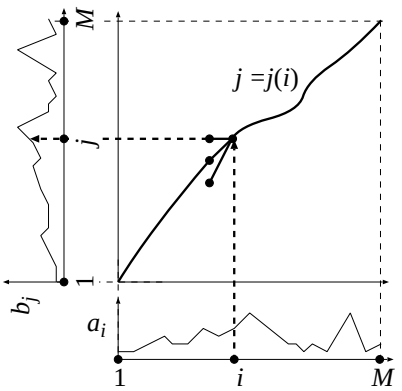

図2　単調連続性制約下でのライン間の画素対応付け関数 $j = j(i)$.

ついては次の検討事項とするが，本報告でもその基本的方針については触れることとする．

本手法の性能評価のために，標準動画像の可逆圧縮実験を行う．非可逆圧縮でなく可逆圧縮を選択したのは，本手法の基本的な性能を把握するためである．本手法自体は動き補償法の一種であるから可逆/非可逆の両方の符号化方式において利用できる．

2. 画素を単位とした動き補償

2.1 方式の概要

前述のように本報告では画素単位動き補償法に関する検討の第一段階として，水平方向に限定した場合について検討する．この場合の動き補償は，連続する2フレームの同一ラインの間の1次元的な画素対応決定問題となる．

第tおよび$t-1$フレームの第lライン$(1 \leqq l \leqq L)$を

$$\boldsymbol{a} = a_1, a_2, \ldots, a_i, \ldots, a_M$$

$$\boldsymbol{b} = b_1, b_2, \ldots, b_j, \ldots, b_M$$

とそれぞれ表す(図1)．ここでa_iおよびb_jはそれぞれRGB成分よりなる3次元ベクトルとする．水平方向の動き補償のために，図1(a)に示すように，これら2ラインの間に，画素を単位として対応付けを行う．これを関数$j = j(i)$として表す．関数$j(i)$により画素a_iは画素$b_{j(i)}$に対応づけられる．この$j(i)$による動き補償の結果，画素a_iの予測残差は

$$\Delta a_i = a_i - b_{j(i)} \tag{1}$$

となる．

動き補償の目的は予測残差の極小化にあるので，$j(i)$の最適化の基準を以下のように定める．

$$\sum_{i=1}^{M} \left\| a_i - b_{j(i)} \right\| \ \rightarrow \ \text{minimize w.r.t. } j(i) \tag{2}$$

ここで記号$\|\cdot\|$は，3次元ベクトルのL1ノルムとする．可逆符号化を行うためには，対応付け関数$j(i)$を上式の意味で最適化した後，Δa_iと$\Delta j(i)$を可変長符号化したものをすべてのiについて伝送すればよい．ここで $\Delta j(i) = j(i) - j(i-1)$である．受信側ではこれらを復号したものと，復元済の第$l-1$ラインの情報を用いることで，第lラインを完全に復元できる．画像全体を可逆符号化するためには，同様の処理をラインを変えながらL回繰り返すことになる．

2.2 動きの制約

本手法においては，動き情報の符号量を低減するために，動きに対し，次の単調連続性制約を課する．

$$0 \leqq \Delta j(i) \leqq 2 \tag{3}$$

この制約の下では関数$j(i)$は図2に示すように$i-j$平面上の(近似的な)単調連続関数として表現できる．動き差分値$\Delta j(i)$は0,1,2のいずれかとなり，単純に考えても1画素あたり高々2ビットの符号量で済む．動き情報符号化の詳細ついては，**2.4**で述べる．

2.3 動き推定のアルゴリズム

2ライン$\boldsymbol{a}$と$\boldsymbol{b}$間の動き推定問題，すなわち目的関数(2)の制約条件(3)付き最小化問題は，1次元パターン$\boldsymbol{a}, \boldsymbol{b}$のマッチング問題であり，音声認識における時間軸非線形伸縮問題[3]と同様，動的計画法(dynamic programming : DP)により効率的に解くことができる．最適な動きの差分値$\Delta j(i)$および対応する予測残差Δa_iを求めるためのDPアルゴリズムを図3に示す．ここで$g(i,j)$および$p(i,j)$はワークエリアである．ステップ4のいわゆるDP漸化式を計算していくことで，最終的に$g(M,M)$として目的関数(2)の最小値が得られる．なおステップ3にあるWは動きを$[-W,W]$に制限するためのパラメータである．ステップ8以降はバックトラック処理と呼ばれるもので，計算済のワークエリア$p(i,j)$を用いて，$i = M$から1まで遡りながら，$j(i)$，$\Delta j(i)$，Δa_iを求めている．

2.4 動き情報の符号化

動き差分$\Delta j(i) \in \{0,1,2\}$の可変長符号化は，各ライン毎に(i) $\Delta j(1), \ldots, \Delta j(M-1)$について値$0,1,2$の頻度を数え，(ii)最頻値に応じて符号表(表1)を切り替えながら実行される．そ

```
    /* Initialization */
1   g(1,1) := ‖a_1 − b_1‖

    /* DP-Recursion */
2   for i := 2 to M do begin
3     for j := i − W to i + W do begin
4       g(i,j) := ‖a_i − b_j‖ + min_{p=0,1,2} g(i − 1, j − p)
5       p(i,j) := arg min_{p=0,1,2} g(i − 1, j − p)
6     end
7   end

    /* Backtrack operation */
8   j := M
9   for i := M downto 2 do begin
10    Δa_i := a_i − b_j
11    Δj(i − 1) := p(i,j)
12    j := j − Δj(i − 1)
13  end
14  Δa_1 := a_1 − b_1
```

図3　動き推定アルゴリズム．

表1　動き差分情報 $\Delta j(i)$ に対する可変長符号の例．

$\Delta j(i)$ の最頻値	$\Delta j(i)$ の値 → 符号		
0	0 → 0	1 → 10	2 → 11
1	0 → 11	1 → 0	2 → 10
2	0 → 10	1 → 11	2 → 0

の結果，各ラインの動き情報として伝送されるのは，この符号表に従って符号化された $(M-1)$ 個の動き差分情報および符号表切り替えのための2ビット (=$\Delta j(i)$ の最頻値を表現) となる．ここで前者が高々 $M/3\times 1+(M-M/3)\times 2=5M/3$ ビットであり，また $M\gg 2$ であることを考慮すれば，動き差分情報は高々 $(5M/3+2)/M\sim 1.67$ bpp (bit per pixel) となる．

3. 実　　験

3.1 使用データ

評価実験においては，標準的に用いられている11種の動画像を用いた(注1)．表2に，それらの名称やサイズ，フレーム数，動きの程度の目安 $(0\sim+++)$，および第1，中央，最終フレームを示す．いずれの動画像もシーンチェンジは含まない．また，各フレームはRGB 各8ビットの3プレーンより成り，従って各画素の原符号量は24ビットとなる．

3.2 使用した動き補償法

本報告では，単純フレーム間差分法および次の3つ動き補償法 (いずれも水平方向の動き補償) を利用し，可逆圧縮実験を通してそれらの比較評価を行った．

(注1): "MomDaughter" 以外は 米国レンスラー工科大学(Rensselaer Polytechnic Institute)のページ http://www.cipr.rpi.edu/resource/sequences/より入手した．

表2　実験で用いた動画像 (グレー画像化して表示)．

name size frms/motion	first, mid, and last frames
Garden 352x240 115 / ++	
Coastguard 352x288 300 / +++	
Foreman 352x288 300 / ++	
Football 352x240 125 / +++	
Mobile 352x240 140 / ++	
Carphone 176x144 96 / ++	
MissAmerica 360x288 150 / +	
Claire 360x288 168 / +	
MomDaughter 360x240 150 / +	
Hall_Monitor 352x288 300 / +	
Container 352x288 300 / ∼ 0	

MC1:　制約付き画素単位動き補償 (本手法/図1(a))

MC2:　無制約-セグメント単位動き補償 (図1(b))

MC3:　無制約-画素単位動き補償 (図1(c))

MC2は，ライン $\boldsymbol{a}$ を16画素長の線分 (セグメント) に分割し，各線分毎に独立に $\boldsymbol{b}$ 中に対応領域を探索することで動き補償を行う方式である．これはブロック/パッチといった小領域を単位とする動き補償法を，ライン上で模したものと言える．MC3は，本手法から単調連続性制約を除いたものであり，各画素独立に対応画素を探索することで動き補償を行う．これらMC2, MC3については，動きの許容範囲が $[-W,W]$ に制限されているものとする．本節および次節**3.3**にて後述するように，このパラメータ W は，計算量および符号化効率に影響する．

表3　各手法の1フレームあたりの時間計算量．
($L \times M$：フレームサイズ (単位:pixel);　W：許容する最大動き (単位:pixel))

	MC1: present MC	MC2: unconstrained segment-based MC	MC3: unconstrained pixel-based MC
horizontal motion only	$O(LMW)$	$O(LMW)$	$O(LMW)$
two-dimensional motion	$O(LMW)$ or $O(LMW^2)$	$O(LMW^2)$	$O(LMW^2)$

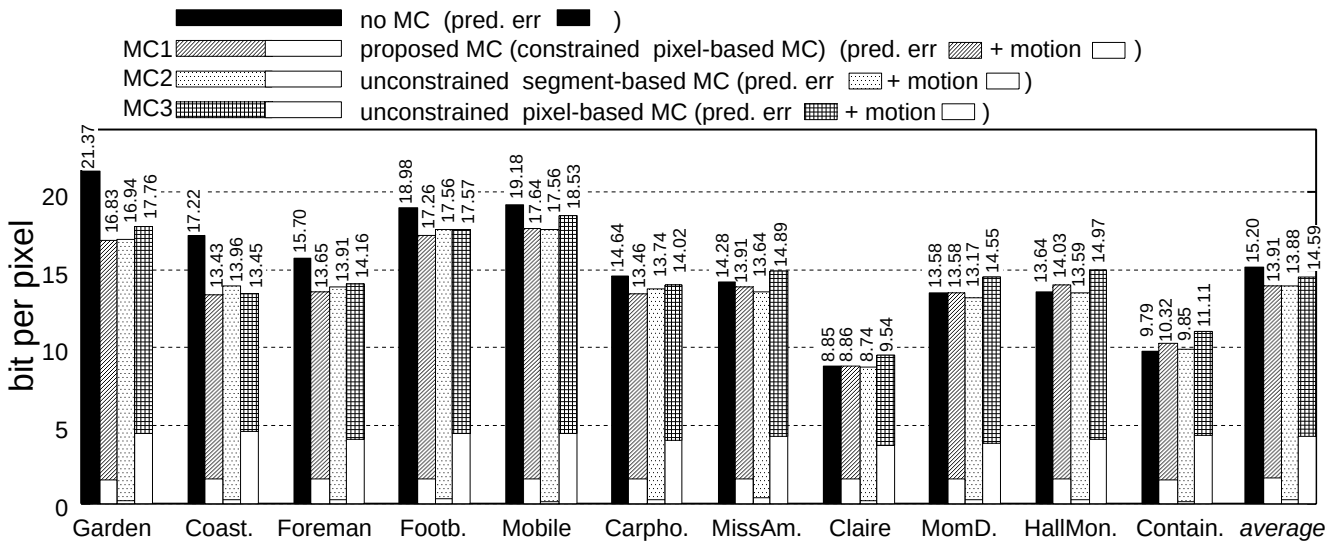

図4　各手法による1画素当りの符号量 (全フレームでの平均値)．

実験においては，標準符号化方式で通常採用されている M-C/no MC判定は使用しなかった．すなわち，前述の動き補償法 (MC1∼ 3) においては，全フレームの全ラインにおいて動き補償を行うこととした．

MC1∼3それぞれについて1フレーム全体を符号化するために要する時間計算量を表3にまとめた．このように，水平方向の動き補償に限定して行う場合，これらMC1∼3の計算量は等しく$O(LMW)$となる．

3.3　評　価　法

各々の圧縮効率を1画素当りの平均符号量で評価した. この平均符号量は，予測残差符号量と動き情報符号量を各フレームで求め，それらを全フレーム分合計したものをフレーム数と1フレーム当たりの画素数で割って求めた． 予測残差符号量と動き情報符号量の算出の詳細は以下の通りである．

(A) 予測残差符号量：まず各フレーム毎に独立にハフマン符号化を行い，全画素分の符号量を実際に算定した．次にハフマン木のための符号量 (予測残差のヒストグラムの符号量) を求めた．両者の合計を1フレーム当りの予測残差符号量とした．

(B) 動き情報符号量：本手法 MC1 の動き情報符号量については，**2. 4**で述べた符号化方式に基づいて算定した．MC2 および MC3 については以下の通り算定した．まず隣接画素もしくはセグメント間で動き差分を求めた．この差分の範囲は$[-2W, 2W]$となるが，これは損失なく$[-W, W]$の範囲に変換できる [2]．その後，ハフマン符号化を行い，その符号量にハフマン木に要する符号量を加算したものを1フレーム当りの動き情報符号量とした．

MC2およびMC3については，動きの許容範囲$[-W, W]$に応じて符号化効率が変化する．一般には，許容範囲を狭くすると，予測残差符号量は増え，動き情報符号量は減少する．許容範囲を広くするとその逆となる．以下の実験結果においては，パラメータWを16から$M/4$まで変化させながら総符号量が最も小さくなった場合を示した．なお，Wによる総符号量の変化幅はそう大きくはなく，MC2, MC3においてそれぞれ高々0.2 bpp, 1 bpp 程度であった．本手法 MC1 にもパラメータWはあるが (図3 ステップ3)，動き情報符号量にはほとんど影響しない．よって，符号量最小化の目的からは予測残差量を最小化できる$W = \infty$と設定すべきであり，実験においてもそれに従った．

3.4　結果と考察

以上の実験の結果を図4に示す．同図では，左から右に向って，本手法 MC1 の単純フレーム間差分法に対する改善幅が大きかったものから順に，11種の動画像に対する結果を並べてある．また右端はそれらの平均値である．MC1∼ 3については，符号量を動き情報分と予測残差分に分けて示してある． この結果から以下のように考察できる．

(1) 単純フレーム間差分法に対し，本手法により平均して1.3 bpp，最高で4.5 bppの改善が見られた．厳密に言えば，動きの少い動画像 ("Container" など) についてはわずかながら本手法のほうが符号量が多くなっている (最悪で 0.5 bpp 程度増加)．ただしこうした増加はこれは他の動き補償法 (MC2, MC3) にも同様に見られた．

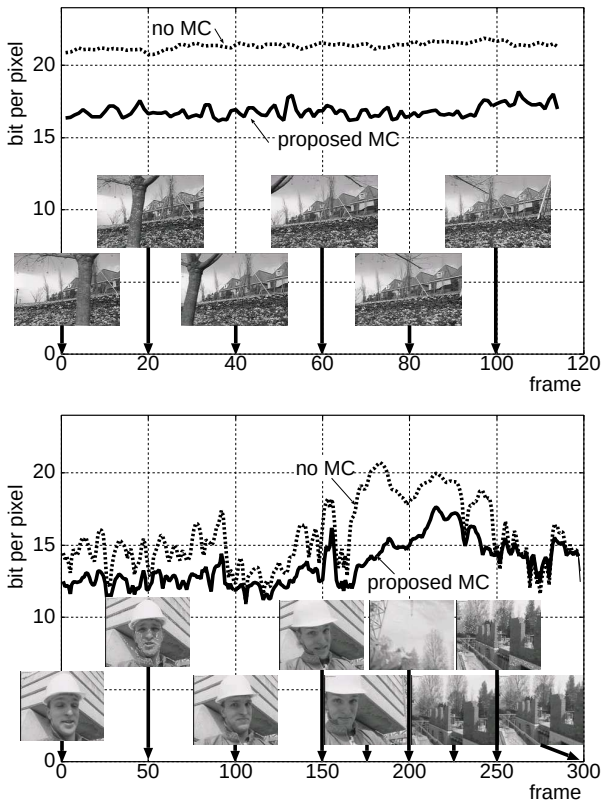

図5　フレーム毎の1画素当り符号量の変化．(上)Garden, (下) Foreman．

（2） 平均的には本手法 MC1 と セグメント単位の手法 MC2 とはほとんど同等の符号化性能であった．ただし両者をより詳細に見ると，動きの大きな動画像についてはMC2に比べて本手法のほうが優位であった．このことは，従来より広く用いられている小領域単位の動き補償法よりも，画素単位の動き補償法の方が高圧縮率を達成しうることを示唆している．反面，動きの小さな動画像については，MC2のほうが優位となっている．

（3） 本手法の動き情報符号量は，動画像の内容に依らずほぼ一定値(約1.6bpp)であった．

（4） 予測残差符号量が少いという意味ではMC3が最良の方法であった．反面，動き情報符号量が非常に多く，結果的には本手法のほうが常に優位となった．本手法とMC3は，同じく画素を単位とした動き補償法であるが，このように動き情報符号量の間に大きな差が生じたのは，本手法における単調連続制約が効果的であったことを示している．

以上を総合すると，本手法は特に動きの大きな動画像に対して，有効な手法であると言える．

図5に，本手法と単純フレーム間差分法について，フレーム毎の符号量の変化を示す．この図から，本手法の単純フレーム間差分法に対する安定した優位性がわかる．この優位性は”Garden”の全フレーム，および “Foreman” の 150∼250 フレームにおいて特に顕著に認められる．これら部分ではカメラパンニングにより大きな動きが発生しており，従ってこの結果は本手法の動き補償能力を示すものとなっている．なお，図には示さなかったが，MC2とMC3の符号量の変化の様子は本手法とほぼ同様であった．

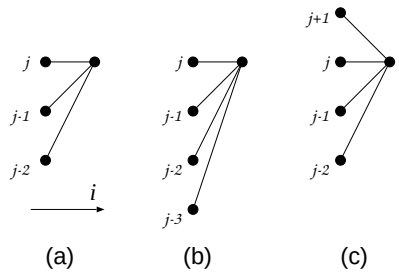

図6　対応付け関数$j = j(i)$の局所傾き．

4. 検　　討

本節では，本手法と既存技術との関係を概観しながら，本手法に関する将来的な検討課題を中心として議論する．

4.1 他の技術との関連

4.1.1 ステレオマッチングとの関連

2ライン間の画素対応付けを行う本手法のアルゴリズムは，エピポーラ拘束下でのステレオマッチングアルゴリズム [4], [5] の一種と見なすことができる．このことは，ステレオマッチングアルゴリズムにおける様々な考え方を本手法に組み込めること示唆している．例えば，ステレオマッチングにおいては，オクルージョンに対応するため，非単調/不連続な対応付け関数が利用されることがある．この考え方を本手法に組み込むとすれば，対応付け関数$j(i)$の局所傾きを，現在の図6(a) から，同図 (b) や (c) のように変更すればよい．また，ステレオマッチングの1 形態である区間対応法 [6] の考え方は，動き情報の削減につながる可能性もある．

4.1.2 パターン認識における弾性マッチング法との関連

2画像間の画素対応関係を最適決定する手法は，パターン認識，特に文字認識のための手法としても，弾性マッチングや非線形伸縮マッチングといった名称で1970年代から様々に提案されており，本手法を改良/拡張する上で参考となり得る．実際，次節**4.2**において述べる動きの2次元化においては， それら従来法の考え方を積極的に利用する予定である．

弾性マッチング法においてはパターンの2次元的位相構造を保つための制約が課されていることが多い．一方，位相構造を保つ必要のない動き補償においては，こうした制約が利用されることはなかった．しかし，実験において本手法 MC1 がMC3 に対して優位であったように，動き情報符号量を低減できるという意味で，この制約が動き補償にも有効である可能性がある．

4.1.3 静止画像圧縮技術との関連

ライン$\boldsymbol{a}, \boldsymbol{b}$を同一フレーム内の隣接ライン (第$l$ラインおよび第$l-1$ライン) と見なせば，本手法はそのままフレーム内予測符号化手法としても利用することができる．この場合，本手法はフレーム内予測符号化の一手法であるライン予測 [1] の拡張形に相当する．

4.2 動きの2次元化

本報告では，水平方向の動き補償に限定して議論を進めてき

た．しかしながら動きは本来2次元的であるので，次ステップとして，動きの2次元化を検討することは自然である．

以上に述べた枠組みにおいて動きを2次元化することを考えると，制約条件 (3) を垂直方向にも仮定した上で，a_i に最も近い画素を，画像 $\boldsymbol{B}$ において2次元的に探索することになる．しかしこのように制約が2次元的になると，残差最小化問題の大局的最適解を求めることは計算量的に非常に困難な問題となる [7], [8] ．そこで何らかの高速な近似解法が必要になってくるが，パターン認識の分野においてはそうした要求を満たすような手法が既に幾つか提案されており，従ってそれらを動き補償において利用することが第一に考えられる．例としては，本手法を水平方向/垂直方向に2段階的に適用する手法 [9] や，ラインを一括して上下動させながらライン内で1次元的な対応づけを行う手法 [10], [11] が挙げられる．今後はこれらを含め様々な手法を画素単位動き補償に利用し，動き情報の符号量を考慮しながらその性能を確認する必要があると考えている．

MC1∼ 3を2次元的な動き補償を行うように拡張した場合の時間計算量を表3の下段にまとめた．本手法の項の $O(LMW)$ と $O(LMW^2)$ は，画素対応付けにそれぞれ上述の文献 [9] と文献 [10], [11] の方式を用いた場合の計算量である．MC2の2次元拡張形はいわゆるブロックベース動き補償である．この表よりわかるように，画素単位の動き補償法は，ブロックベースの手法よりも少い計算量となりうる．

DPによる2次元的な動き補償法としては，既にChenら [12] の方法がある．方法はメッシュベースの動き補償法として提案されたものであり，画素を単位とすることを出発点とする本手法とは本質的に異なる．ただし参考になる点も多く，例えば動きベクトルの最適化問題の目的関数として予測残差エネルギーに動きベクトルの符号量を加えるなどの工夫は本手法においても利用可能である．

5. ま と め

画素を単位とした動き補償法に関する検討の第一段階として，水平方向に限定した場合について考察を行った．RGB動画像 (24 bpp) の圧縮実験の結果，動き補償を行わない場合，すなわち単純フレーム間差分法に比べ，平均して1.3 bpp，最高で4.5 bppの改善を確認した． また，実験を通して小領域を単位とする動き補償法との比較も行った．その結果，平均的には両者は同程度の圧縮率が得られるが，動きの大きな動画像においては本手法のほうが高い圧縮率を得ることを確認した．

今後はまず本手法を2次元的な動きに対応できるように拡張する予定である．この拡張の基本的指針に関しては既に本報告の**4. 2**においても触れた．また，本報告では可逆圧縮の枠組みで比較評価したが，より利用範囲の広い非可逆圧縮の枠組みにおいても評価すべきと考えている．非可逆圧縮において従来の小領域を単位とする動き補償を利用した場合，その小領域内での動きの均質性に起因して，いわゆるブロック歪みなど，種々の歪みが発生することが知られている [13]． 画素を単位することで，こうした歪みを回避できるのではないかと考えている．

文 献

[1] 原島 (監), 画像情報圧縮, オーム社, 1991.

[2] 安田, 渡辺, ディジタル画像圧縮の基礎, 日経BP出版センター, 1996.

[3] 迫江, 千葉, “動的計画法を利用した音声の時間正規化に基づく連続音声認識,” 日本音響学会誌, vol. 27, no. 9, pp. 483-490, 1971.

[4] 磯道, 竹政, “航空写真からの高度図の作製,” 第7回画像工学コンファレンス, pp. 31–34, 1976.

[5] I. J. Cox, S. L. Hingorani, and S. B. Rao, “A maximum likelihood stereo algorithm,” Comput. Vis. Image Und., vol. 63, no. 3, pp. 542–567, 1996.

[6] 大田, 正井, 池田, “動的計画法によるステレオ画像の区間対応法,” 信学論, vol. J68-D, no. 4, pp. 554–561, 1985.

[7] 内田, 迫江, “動的計画法に基づく単調連続2次元ワープ法の検討,” 信学論, vol. J81-D-II, no. 6, pp. 1251-1258, 1998.

[8] S. Uchida and H. Sakoe, “An Efficient Two-Dimensional Warping Algorithm,” IEICE Trans. Info. & Syst., vol. E82-D, no. 3, pp. 693–700, 1999.

[9] 磯道, 小川, “動的計画法によるパターン・マッチング,” 情報処理, vol. 16, no. 1, pp. 15–22, 1975.

[10] 三輪, 小原, 牧野, 城戸, “非線形スペクトルマッチングによる単語音声認識の一方式,” 信学論, vol. J64-D, no. 1, pp. 46-53, 1981.

[11] S. Kuo and O. E. Agazzi, “Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models,” IEEE Trans. PAMI, vol. 16, no. 8, pp. 842–848, 1994.

[12] M. C. Chen and A. N. Willson, Jr., “Motion-vector optimization of control grid interpolation and block motion compensation using iterated dynamic programming,” IEEE Trans. Image Processing, vol. 9. no. 7, pp. 1145–1157, 2000.

[13] M. Yuen and H. R. Wu, “A survey of hybrid MC/DPCM/DCT video coding distortions,” Signal Processing, vol. 70, no. 3, pp.247–278, 1998.