

Matching Pursuits における斜めパターン対応辞書と学習による設計

加古 真哉[†] 今村 幸祐^{††} 橋本 秀雄^{††}

† 金沢大学大学院自然科学研究科 〒 920-8667 石川県金沢市小立野 2-40-20

E-mail: †kako@gin.ec.t.kanazawa-u.ac.jp, ††{imamura,hasimoto}@is.t.kanazawa-u.ac.jp

あらまし 動画の符号化は動き補償予測 (MC) と離散コサイン変換 (DCT) を組み合わせたハイブリット方式が主流であるが、低ビットレートで有効な動画符号化方式として波形符号化法 Matching Pursuits (MP) を用いる手法がある。MP は基底ベクトルの集合である辞書により信号を近似的に再構成する手法である。MP では計算量を低減させるために、一次元に分離可能な辞書を用いる。しかし分離可能辞書は斜めパターンに対して有効な基底を作成できない。そこで座標変換による斜めパターンに対応した辞書の作成を行う。また、ベクトル量子化のコードブックデザインを参考とした、学習による辞書の設計法についても提案する。

キーワード 動画符号化, Matching Pursuits, 辞書設計, ベクトル量子化

Matching Pursuits using Slant Patterns
and Dictionary Design by learningShinya KAKO[†], Kousuke IMAMURA^{††}, and Hideo HASHIMOTO^{††}

† Graduate School of Natural Science & Technology, Kanazawa University

2-40-20, Kodatsuno, Kanazawa-shi, Ishikawa, 920-8667, Japan

E-mail: †kako@gin.ec.t.kanazawa-u.ac.jp, ††{imamura,hasimoto}@is.t.kanazawa-u.ac.jp

Abstract Video coding methods have been based on hybrid coding which employs motion compensation and discrete cosine transform. But, another high efficiency video coding method in low bit rate using matching pursuits, that is one of waveform coding techniques, is suggested recently. The matching pursuits algorithm represents a signal approximately using a dictionary. Matching pursuits generally use separable dictionary composed of one dimensional basis functions. But, it is difficult for separable dictionary to code images including slant patterns efficiently. In this paper, we design the dictionary for slant patterns by a simple geometric transform. And we propose a dictionary design algorithm by employing learning technique like codebook design of vector quantization.

Key words Video Coding, Matching Pursuits, Dictionary Design, Vector Quantization

1. はじめに

MPEG や H.26X 等の動画符号化方式は動き補償 (MC) と離散コサイン変換 (DCT) による波形符号化を組み合わせたハイブリット方式をベースとしている。これらの方式は MC で予測画像を作成し、DCT により予測誤差画像を符号化することで情報量の削減を行う。Matching Pursuits (MP) は Mallat, Zhang により提案された波形符号化アルゴリズムである [1]。DCT を用いる代わりに MP により予測誤差画像を符号化する動画符号化方式が R.Neff, Zakhor により提案されている。[2]

MP は基底ベクトルの集合である辞書を用いて符号化対象信号を近似的に再構成する手法である。つまり MP の符号化性能は辞書に大きく依存する。MP で用いられる辞書は、計算量を

少なくするために一次元に分離可能な辞書を用いる。しかし、分離可能な辞書では斜め方向の基底を構成できないため、予測誤差画像の斜めパターンを表現する際に非効率となる問題がある。また、辞書は三つの関数パラメータを持つ Gabor 関数により作成されており、このパラメータを最適化することで符号化効率の改善が見込まれる。

本稿では、辞書ブロックの座標変換による斜めパターンに対応した辞書の設計について提案する。一般的な一次元分離可能辞書を作成後に簡単な座標変換を行う手法であるため、辞書の一次元分離可能という利点を保ちつつ、辞書を斜めパターンに対応させることができる。また、ベクトル量子化のコードブック設計を参考にした、学習による最適な辞書パラメータ決定方法についても提案を行う。

本稿では二章でMPの概略について説明し、三章で今回提案する斜めパターン対応辞書及び学習による辞書設計法について述べる。四章に計算機によるシミュレーション結果、最後に五章で本稿のまとめを記す。

2. Matching Pursuits

2.1 MP符号化の概要

Matching Pursuits(MP)は非直交かつovercompleteな辞書Dに登録された基底ベクトル $g_\gamma(t)$ をシフトした波形の線形結合によって符号化対象信号 $f(t)$ を再構成する手法である。

$$\hat{f}(t) = \sum_{k=1}^m p_k \cdot g_{\gamma_k}(t - \tau_k) \quad (1)$$

ここで τ_k, γ_k, p_k はそれぞれ基底ベクトルの位置、種類、係数を表すパラメータであり、これらの3種類のパラメータの組 (τ_k, γ_k, p_k) をアトム(atom)と呼ぶ。MPでは式(1)を構成するアトムを信号 $f(t)$ の基底ベクトル $g_\gamma(t)$ への直交射影による連続的な処理により決定していく。

まず、信号 $f(t)$ をある基底ベクトル $g_{\gamma_1}(t - \tau_1)$ へ射影することにより信号 $f(t)$ は

$$f(t) = p_1 \cdot g_{\gamma_1}(t - \tau_1) + Rf(t) \quad (2)$$

$$p_1 = \langle f(t), g_{\gamma_1}(t - \tau_1) \rangle \quad (3)$$

と表せる。ここで p_1 は信号と基底ベクトルの内積値であり、 $Rf(t)$ は基底ベクトル $g_{\gamma_1}(t - \tau_1)$ へ射影した時の残差信号である。また基底ベクトル $g_\gamma(t)$ のノルムが1に正規化されているものとする、式(2)において $Rf(t)$ と $g_{\gamma_1}(t - \tau_1)$ は互いに直交していることから、以下の関係式が成立する。

$$\|f(t)\|^2 = p_1^2 + \|Rf(t)\|^2 \quad (4)$$

すなわち内積値 p の絶対値を最大とするパラメータを求めることにより、反復の各段階において残差信号のエネルギーの減少量 $\Delta e = \|f(t)\|^2 - \|Rf(t)\|^2$ を最大となるアトムを決定することができるため、効率良く符号化を行うことができる。

同様に残差信号 $Rf(t)$ を展開すると

$$Rf(t) = p_2 \cdot g_{\gamma_2}(t - \tau_2) + R^2f(t) \quad (5)$$

となる。このような繰り返し処理により m 個のアトムにより展開された信号 $f(t)$ は

$$f(t) = \sum_{k=1}^m p_k \cdot g_{\gamma_k}(t - \tau_k) + R^m f(t) \quad (6)$$

となる。これより m 個のアトムにより信号 $f(t)$ を近似的に再構成することができる。

2.2 MC+MP 動画像符号化法

MPを用いた動画像符号化システム(MC+MP方式)のブロック図を図1に示す。これは現フレームと前フレーム間でMCにより予測フレームを作成し、予測誤差フレームをMPにより符号化を行う。

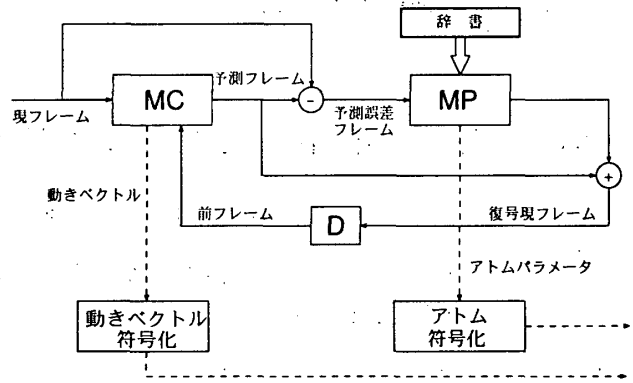


図1 MC+MP符号化システム

MPで用いる辞書はガボール関数(式(8))から生成される2次元分離可能な辞書である。

$$G_{\alpha, \beta}(i, j) = g_{\alpha}(i) \otimes g_{\beta}(j) \quad (7)$$

\otimes : 直積を表す

$$g_\gamma(i) = K_\gamma g\left(\frac{i - \frac{N}{2} + 1}{s}\right) \cos\left(\frac{2\pi\xi(i - \frac{N}{2} + 1)}{N} + \phi\right) \quad (8)$$

ここで、 $g(\cdot)$ はガウシアンウィンドウであり、 K_α は正規化係数である。このガボール関数は $\gamma = (s, \xi, \phi)$ により特定される。これらのパラメータと式(7)により作成した辞書を図2に示す。

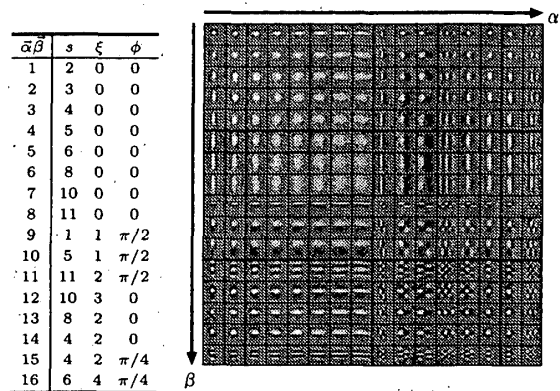


図2 基本辞書

3. 辞書の設計

MPは辞書により符号化対象画像を近似構成するため、符号化特性は辞書に大きく依存する。つまり、より符号化対象画像に適した効率の良い辞書の設計により、符号化性能の改善が見込まれる。本稿では辞書の座標変換による斜め方向の画像パターンに対応した辞書の構成方法を示す。この手法は辞書の一次元分離可能という利点を保ちつつ、斜めパターンへの対応が可能な辞書の設計法である。次に学習による辞書パラメータの最適化方法について示し、より効率の良い辞書の設計を行う。

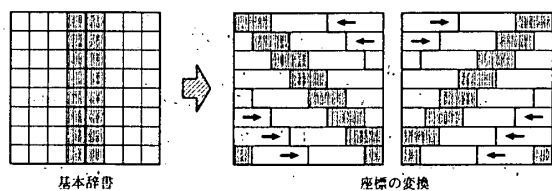


図3 辞書の座標変換

3.1. 斜めパターンへの対応

MPの辞書は一次元に分離可能な辞書を用いるのが一般的であり、この特徴により最適な辞書ブロックの探索計算量を削減している。しかし、式(7)で表される2つの1次元Gabor関数の直積により作成された辞書(図2)により符号化する従来法では、斜めパターンの構成は困難であり、斜めが多く現れる画像を効率的に符号化できない。

そこで本稿では座標変換による斜めパターンの作成を提案し、その方法について示す。図3に示すように、辞書を水平方向に階段状にずらすことによって、基本辞書で直線の部分が斜めのパターンとして表され、斜め方向のエッジに対応する辞書が作成出来る。

図2の基本辞書を座標変換した辞書を図4に示す。これらの処理は単純な座標変換のみで構成されるため、分離可能な特性を維持することができる。ここでは、これらの辞書をシフト辞書と呼ぶものし、右にずらした辞書(図4(a))を右シフト辞書、左にずらした辞書(図4(b))を左シフト辞書とする。また、3種類の辞書のうちどの辞書から選択されたかを示すインデックスを ω と定義する。

式(9)に二次元シフト辞書の作成式、式(10)に右シフト辞書、(11)に左シフト辞書の座標変換式、(12)に座標補正値 k を示す。

$$G_{\alpha,\beta}(i,j) = g_{\alpha}(i+k) \otimes g_{\beta}(j) \tag{9}$$

$$I_R = i + j - \left(\frac{N}{2} - 1\right) \tag{10}$$

$$I_L = i - j + \left(\frac{N}{2} - 1\right) \tag{11}$$

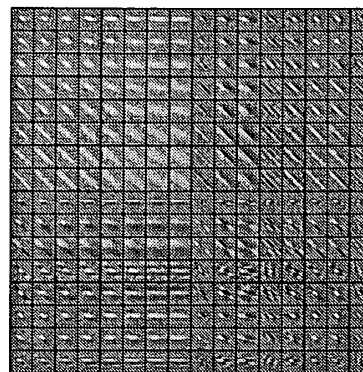
$$k = \begin{cases} 0 & (0 \leq I \leq N) \\ N & (I < 0) \\ -N & (N < I) \end{cases} \tag{12}$$

シフト辞書を用いた動画像符号化システム(以後シフト辞書手法とする)のATOM決定手順を示す。

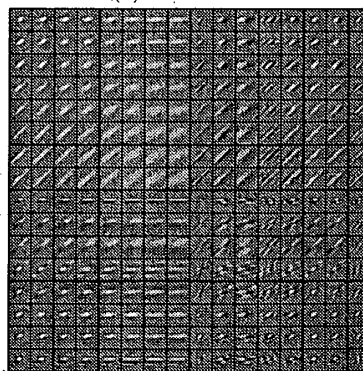
1. 基本辞書(図2)、右シフト辞書、左シフト辞書の3辞書の各ブロックごとの内積値を計算し、最大内積となるブロックを決定
2. 決定した3辞書の最大内積値を比較し、内積値が最も大きいブロックを選択

3.2 学習による辞書設計

MPで用いる辞書は3つの関数パラメータをもつGabor関数により作成されており、一般に最適な辞書設計を行うことは



(a) 右シフト辞書



(b) 左シフト辞書

図4 シフト辞書

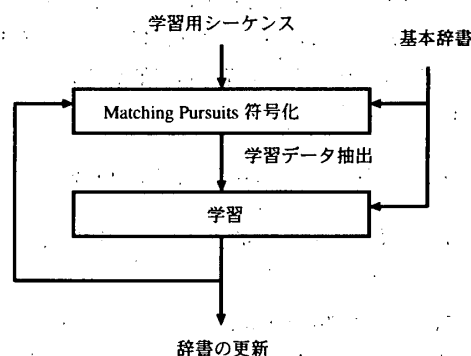


図5 辞書学習アルゴリズム

困難である。そのため、辞書設計のための検討も行われている。[5]

本稿では、辞書の学習による最適辞書設計を提案する。辞書設計アルゴリズムの概略を図5に示す。

• 学習アルゴリズム

本稿で提案する辞書学習の基本アルゴリズムを以下に示す。

1. 学習用シーケンスに対して Matching Pursuits 符号化を行い学習のためのATOM情報を得る。ここでATOM情報とは、適用した辞書インデックスと適用されたブロックの画素情報である。
2. 2次元の辞書を構成する2つのGabor関数のうち、まず垂直成分 g_{β} を固定し、水平成分 g_{α} のパラメータ α を評価式が最大となるように更新する。本稿では評価は式(13)に示すように、ブロックデータと更新パラメータによる辞書との内積

の総和を用いる。

$$p = \sum_{k=1}^m \langle a_k, g_\alpha \otimes g_\beta \rangle \quad (13)$$

ここで、 a は、更新された α または、 β によって変化する辞書が適用されたブロックの画素情報を示し、 m は辞書が適用されたブロックの数を表す。

3. 2. と同様に、2つの Gabor 関数のうち、水平成分 g_α を固定し、垂直成分 g_β のパラメータの更新を行う。

4. パラメータの変化が収束していない場合、1. に戻り、収束するか、最大繰り返し回数まで繰り返す。

●使用メモリ削減

この辞書学習アルゴリズムのトレーニングデータとなるアトム情報には、符号化対象となったブロックの画素情報が含まれる。学習には多くのトレーニングデータが必要となるが、それをそれぞれ記録しておくためには、メモリが大量に必要となる。これを改善するために、本稿では辞書が適用されたブロックの画素情報を、それぞれ記録するのではなく、データの合計値を代表データとして記録し、それによって学習を行う手法についても検討する。通常学習に必要なメモリ量 M_n は式 (14)、データ合計値を用いることによるメモリの使用量 M_s は式 (15) で表される。通常必要なメモリ量に比べ、データ合計値によるメモリ量は数%以下に抑えることができる。また、通常、フレーム数、アトム数など学習に用いる情報が増加すると必要なメモリ数も増加するが、データ合計値では必要なメモリが変動しない利点もある。

$$M_n = (\text{フレーム数}) \times (\text{アトム数}) \times (\text{辞書ブロックサイズ}) \quad (14)$$

$$M_s = (\text{辞書ブロック数}) \times (\text{辞書ブロックサイズ}) \quad (15)$$

●シフト辞書手法を用いた学習

3.1 で提案したシフト辞書手法と 3.2 で提案した学習による辞書の最適化を組み合わせ、更なる符号化効率の改善を図る。学習を行う際に基本辞書だけでなくシフト辞書も最適化を行う。学習データを抽出する際 (図 5 の Matching Pursuits 符号化) にシフト辞書手法を適用し、学習データを得る。その学習データを用いて基本辞書、シフト辞書の辞書の最適化を行う。

4. シミュレーション

テスト動画として標準動画像 Akiyo, Mother & Daughter (Mother), Carphone, Foreman (QCIF: Y 信号のみ, 50frames, 10 fps) の 4 種類の動画像を用い、計算機によるシミュレーションを行った。

4.1 斜めパターン対応辞書

斜めパターンに対応したシフト辞書を用い、シミュレーションを行った。シフト辞書手法の符号化特性を図 6、その時の各辞書 (基本辞書, 右シフト辞書, 左シフト辞書) の使用比率を表 1 に示す。

シフト辞書手法は 3 つの辞書からどの辞書を適用したか判別

表 1 辞書使用比率

	akiyo	mother	carphone	foreman
atom 50	50:26:24	49:26:25	46:23:31	38:34:28
atom 100	46:28:26	48:26:26	44:26:30	35:35:30
atom 200	44:28:28	46:27:27	41:27:32	35:35:30
atom 400	40:30:30	42:28:29	40:28:32	35:36:29

(基本辞書: 右シフト辞書: 左シフト辞書)[%]

する情報がアトム情報に付加されるため、その分の情報量が増加する。しかし図 6 から、辞書情報の増加分を含んでも、基本辞書のみを用いた手法より符号化効率が改善されていることがわかる。また、ビットレートの増加にともないシフト辞書手法の効率改善度合いが大きくなっている。表 1 において、シフト辞書はおよそ 50~60% 選択されていることから、シフト辞書の適用が画像の符号化効率改善に貢献していることが示された。

図 7 に従来法と提案法で構成された画像比較例 (foreman) を示す。図から分かるように、提案法により背景の斜め部分がより鮮明に構成されている。

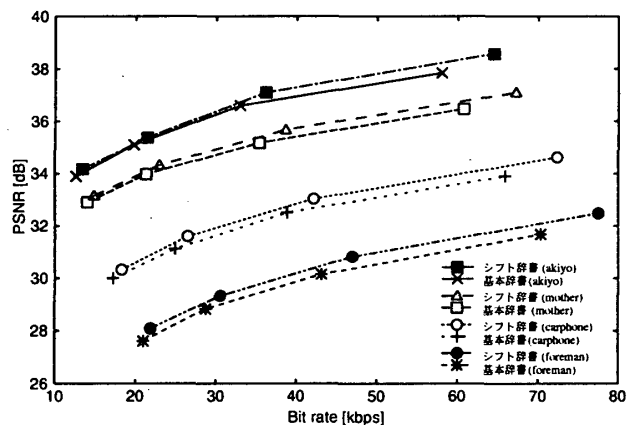


図 6 符号化特性比較

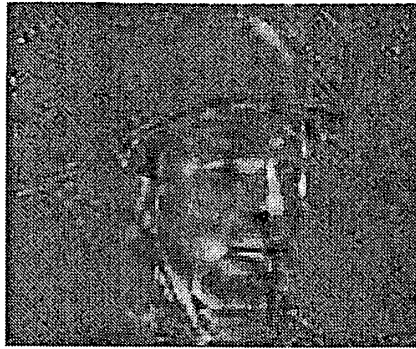
4.2 学習による辞書設計

次に学習による辞書設計法に関して、シミュレーションを行った。

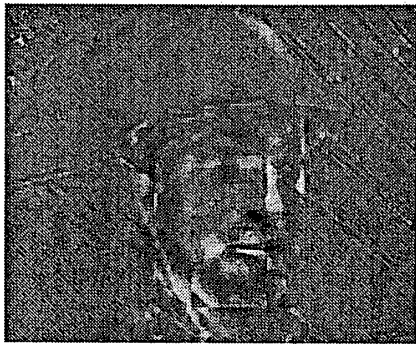
テストシーケンスは、akiyo, carphone, foreman, mother & daughter (50 frames, 10fps) の 4 種類を用い、学習データとなるアトム数は 400 に設定した。学習によって変更する辞書パラメータは、辞書の構成に影響の大きいパラメータである s , ξ とし、 ϕ は固定した。また、パラメータの更新は、 s は、0~16 まで 1 間隔、 ξ は 0~8 まで 1.0 間隔の値について、それぞれ網羅的に評価を行い、式 (13) が最大になるパラメータに更新を行った。学習の繰り返し回数は最大 30 回に設定した。

符号化対象画像に最適な辞書の設計

まず、理想的なトレーニングデータが得られたものとして、符号化対象シーケンスから得られた学習データにより辞書の設計を行った。図 8 にテストシーケンスにおける基本辞書の特性と最適化辞書の符号化特性を示す。図 8 から、学習によって最適化された辞書により、全ての画像で特性が改善されたことが分かる。特に akiyo の改善効果が大きく、約 0.3~1.7dB 向上



(a) 基本辞書



(b) シフト辞書

図 7 再構成画像比較 (foreman, 004 フレーム, アトム数 400)

している。

次にシフト辞書手法を用いた学習により設計を行った場合の特性を示す。図 9 にシフト辞書手法の特性と最適化シフト辞書手法の符号化特性の比較を示す。図から、3.1 章で提案したシフト辞書手法の符号化効率が学習による辞書設計によって改善されている。

汎用的な辞書の設計

符号化対象画像の学習データによる辞書は、他の画像においても最適であるとはいえない。そこでテストシーケンスを含まない 10 種類のシーケンス (boat, bus, bicycle, cheer leader, fest, football, girl, kiel, mobile&calender, tabletennis) から抽出した学習データにより辞書の最適化を行い、汎用的な最適辞書の設計を行った。学習データを抽出したシーケンスは、全て 30 frames, 10fps, で、各フレーム毎に 400 アトム分のアトム情報を抽出した。

図 10 に、基本辞書に対して学習を行った場合の特性を示す。図に示されるように、全ての画像において符号化効率が改善された。テストシーケンスを含まない 10 種類の画像から学習データを得ているため、符号化対象シーケンスから学習データを得た場合に比べ符号化効率の改善度合はやや低下しているが、汎用的な辞書の設計が成されている。

次に、シフト辞書手法に対して学習を行った辞書の特性を図 11 に示す。図 11 は学習を行う前の基本辞書、学習を行う前のシフト辞書手法、との比較を行っている。シフト辞書手法と辞書の学習を組み合わせた手法が最も良い符号化効率を示している。以上の結果から、シフト辞書手法、辞書の学習による最適

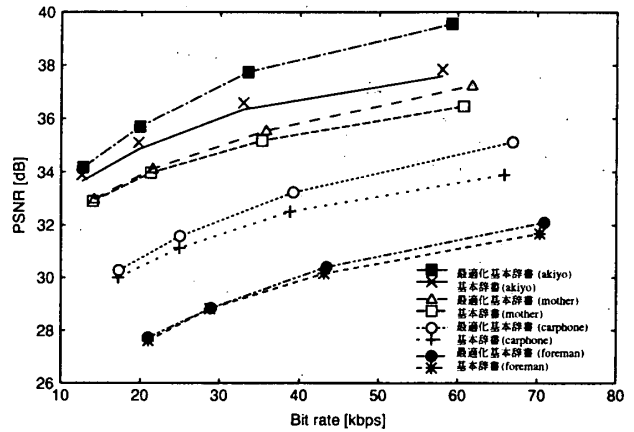


図 8 符号化特性 (基本辞書, 符号化対象シーケンス)

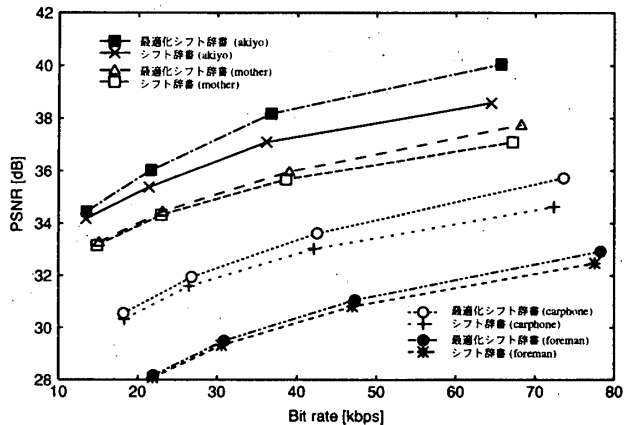


図 9 符号化特性 (シフト辞書, 符号化対象シーケンス)

表 2 最適化辞書パラメータ

(a) 基本辞書						(b) シフト辞書					
	α		β		α, β		α		β		α, β
	s	ξ	s	ξ	ϕ		s	ξ	s	ξ	ϕ
1	4	4	5	5	0	1	3	4	5	5	0
2	4	1	4	1	0	2	4	1	4	1	0
3	16	1	16	1	0	3	15	1	15	1	0
4	5	0	5	0	0	4	5	0	5	0	0
5	6	0	6	0	0	5	6	0	6	0	0
6	8	0	8	0	0	6	8	0	8	0	0
7	12	0	12	0	0	7	10	0	10	0	0
8	0	0	0	0	0	8	16	0	16	0	0
9	4	4	4	1	$\pi/2$	9	4	4	4	4	$\pi/2$
10	5	1	5	1	$\pi/2$	10	5	1	5	1	$\pi/2$
11	16	1	16	3	$\pi/2$	11	16	1	16	1	$\pi/2$
12	10	3	8	2	0	12	9	3	9	3	0
13	9	2	9	2	0	13	9	2	9	2	0
14	4	2	4	2	0	14	4	2	4	2	0
15	4	2	4	2	$\pi/4$	15	4	2	4	2	$\pi/4$
16	6	4	5	4	$\pi/4$	16	6	4	6	4	$\pi/4$

化の効果が示され、また、この 2 手法を組み合わせ、MP の符号化効率をより改善できることが実証された。

表 2 に学習によって得られたパラメータの表を示す。

メモリ使用量削減

学習アルゴリズムにおいて、メモリ使用量の削減のために、各辞書毎に適用されたブロックの合計値を記録して学習に利用する方式の特性を示す。

表 3 に、合計評価法のアトム数における PSNR 特性を示す。これまでの基本辞書、最適データ学習、10 種画像データ学習の特性も同時に示す。表の合計評価法は 10 種類画像データ学習に適用したものである。基本辞書に関しては、合計評価法を用

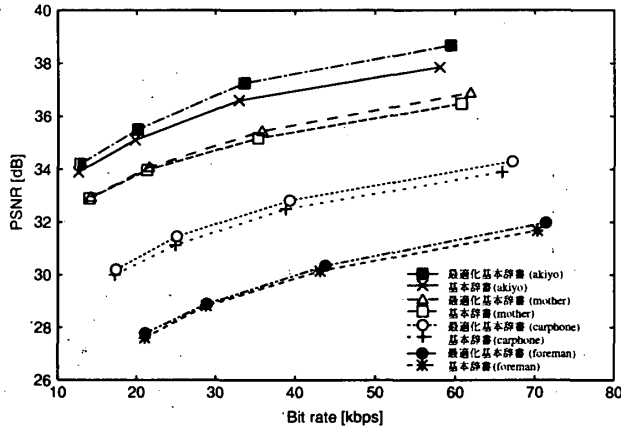


図 10 符号化特性 (基本辞書, 10 種画像シーケンス)

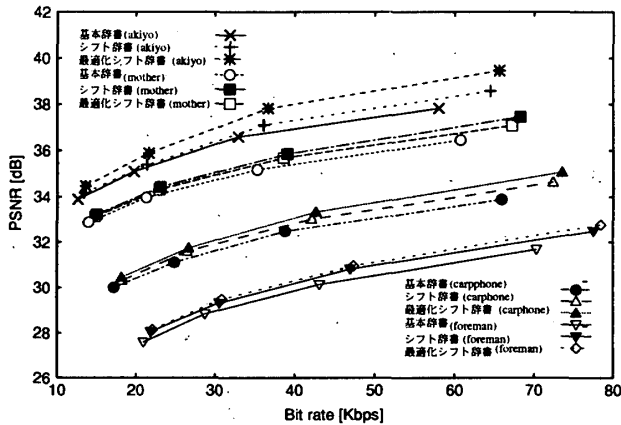


図 11 符号化特性 (シフト辞書, 10 種画像シーケンス)

表 3 PSNR(基本辞書)

Dictionary	atom	PSNR			
		akiyo	mother	carphone	foreman
基本辞書	50	33.90	32.89	29.99	27.60
	100	35.10	33.97	31.12	28.83
	200	36.60	35.18	32.50	30.15
	400	37.84	36.47	33.89	31.68
学習辞書 (符号化対象 シーケンス)	50	34.18	32.97	30.28	27.74
	100	35.70	34.14	31.58	28.85
	200	37.75	35.56	33.22	30.39
	400	39.56	37.26	35.13	32.10
学習辞書 (10 種 シーケンス)	50	34.19	32.94	30.20	27.78
	100	35.50	34.08	31.44	28.89
	200	37.24	35.45	32.81	30.36
	400	38.68	36.87	34.31	31.96
学習辞書 (10 種 シーケンス 合計評価)	50	34.22	32.96	30.16	27.76
	100	35.44	34.10	31.41	28.94
	200	37.15	35.43	32.83	30.34
	400	38.67	36.86	34.40	32.00

いても PSNR 特性の低下はほとんどみられなかった。

5. まとめ

本稿では MP を用いた動画画像符号化の性能改善を目的とし、斜めパターンに対応した辞書の作成および学習によるパラメータ最適化による辞書の設計方法について提案を行った。斜めパ

表 4 PSNR(シフト辞書手法)

Dictionary	atom	PSNR			
		akiyo	mother	carphone	foreman
シフト辞書	50	34.17	33.14	30.33	28.06
	100	35.38	34.32	31.59	29.33
	200	37.11	35.68	33.03	30.82
	400	38.58	37.09	34.62	32.49
学習 シフト辞書 (符号化対象 シーケンス)	50	34.45	33.27	30.55	28.19
	100	36.03	34.45	31.96	29.50
	200	38.19	35.97	33.63	31.06
	400	40.05	37.75	35.71	32.94
学習 シフト辞書 (10 種 シーケンス)	50	34.45	33.24	30.45	28.13
	100	35.89	34.43	31.75	29.47
	200	37.82	35.86	33.35	30.98
	400	39.46	37.45	35.06	32.77
学習シフト辞書 (10 種 シーケンス 合計評価)	50	34.45	33.24	30.45	28.14
	100	35.83	34.42	31.72	29.43
	200	37.64	35.79	33.19	30.95
	400	39.16	37.31	34.86	32.65

ターン対応辞書 (シフト辞書) は座標変換により作成され、一次元に分離可能な特性を保ったまま斜めパターンを構成が可能となる。また、学習による辞書設計ではベクトル量子化のコードブック設計を参考とし、辞書パラメータの最適化を行い辞書を設計し、符号化効率の改善を試みた。斜めパターン対応辞書は基本辞書で構成出来なかった斜めパターンをより正確に構成し、符号化効率も改善した。また、提案した辞書設計法により作成された辞書を用い、符号化効率を向上させることができた。

文 献

- [1] S. Mallat and Z. Zhang, *Matching Pursuits With Time-Frequency Dictionaries*, IEEE Trans. on Signal Processing, Vol.76, No.376, pp.3397-3415, Dec 1993.
- [2] R. Neff and A. Zakhor, *Very Low Bit-Rate Video Coding Based on Matching Pursuits*, IEEE Trans. on Circuits and Systems for Video Technology, Vol.7, No.1, pp.158-171, Feb 1997.
- [3] R. Banham and C. Brailean, *A Selective Update Approach to Matching pursuits Video Coding*, IEEE Trans. on Circuits and Systems for Video Technology, Vol.7, No.1, pp.119-129, Feb 1997.
- [4] I. Matsuda and S. Itoh, *Low Bit-Rate Video Coding Based on Seamless Combination of Motion Compensation and Matching Pursuits*, Proc. of IWAIT 2002, pp.131-136, Jan 2002.
- [5] P. S.-Saugeon and A. Zakhor, *Dictionary Design for Matching Pursuits and Application to Motion-Compensated Video Coding*, IEEE Trans. on Circuits and Systems for Video Technology, Vol.14, No.6, June 2004.
- [6] ITU-T Study Group 15, *Draft Recommendation H.263*, Dec 1995.