

可変ブロックサイズ適応予測を用いた RGB カラー動画の可逆符号化

田嶋 周^{1,a)} 亀田 裕介¹ 松田 一朗¹ 伊東 晋¹

概要: 本稿では、RGB 4:4:4 フォーマットのカラー動画を対象とした高能率な可逆符号化方式を提案する。この方式では、各フレームの各色信号に対して線形予測器を複数用意し、これらを可変サイズのブロック単位で切り替えることで画像の局所的性質に応じた予測を実現している。この線形予測器は、符号化対象色信号だけでなく、動き補償された過去フレームおよび現フレームの符号化済み色信号も参照するように拡張されており、時間方向、空間方向、色信号間の相関を同時に利用可能である。シミュレーションの結果、提案方式は動画符号化の国際標準方式である H.264/AVC の High 4:4:4 プロファイルに基づいた可逆符号化方式と比較して、14%-31%優れた符号化効率を達成することが確認された。

Lossless Coding of RGB Color Video Using Variable Block-Size Adaptive Prediction

Abstract: This paper proposes a lossless coding scheme for color video in RGB 4:4:4 format. The scheme utilizes multiple linear predictors designed for each color signal of each frame to exploit not only spatio correlation, but also temporal and inter-color ones simultaneously. Moreover, they are adaptively selected at each block of variable size. Simulation results indicate that the proposed scheme achieves 14%-31% better coding performance than the lossless coding method based on High 4:4:4 profile of H.264/AVC video coding standard.

1. はじめに

動画符号化の国際標準方式である H.264/AVC[1] や H.265/HEVC[2] は、画像を分割したブロック毎にイントラ予測とインター予測を切り替えることにより、それぞれ、空間方向、時間方向の冗長性を削減している。これに対し著者らは以前、空間方向と時間方向の相関を同時に利用可能な適応線形予測による、モノクローム動画の可逆符号化方式を提案した [3]。この方式では、符号化済み画素群の線形結合により符号化対象画素の予測を行っているが、その際に参照する画素を空間方向だけでなく、動き補償 (MC) された過去フレームにも配置することにより効率的な予測を実現している。

一方、デジタルアーカイブなどの色再現性を重視する分野を中心に、近年 RGB 4:4:4 フォーマットのカラー動画

像の需要が高まっており、高品位な映像データの保存手段として、このフォーマットに適した可逆符号化手法が求められている。しかし、カラー動画は一般に色信号間にも高い相関を有しているため、モノクローム動画を対象として開発された符号化手法をそのまま RGB 動画に適用しても、十分な符号化効率を期待できない。そこで本稿では、RGB カラー静止画像用の可逆符号化方式 [4] を動画信号に拡張し、RGB 4:4:4 フォーマットのカラー動画を効率的に可逆符号化する手法について検討する。また、線形予測器の切り替えを行うブロックサイズを、画像の局所的性質に応じて可変とする手法も導入し、さらなる符号化効率の改善を図る。

2. RGB カラー動画の可変ブロックサイズ 適応予測

2.1 ブロック適応予測

本稿では、RGB 4:4:4 フォーマットのカラー動画をフレーム毎に R→G→B の順で符号化することとする。最初

¹ 東京理科大学 理工学部
Faculty of Science and Technology, Tokyo University of Science

^{a)} taji@itohws01.ee.noda.tus.ac.jp

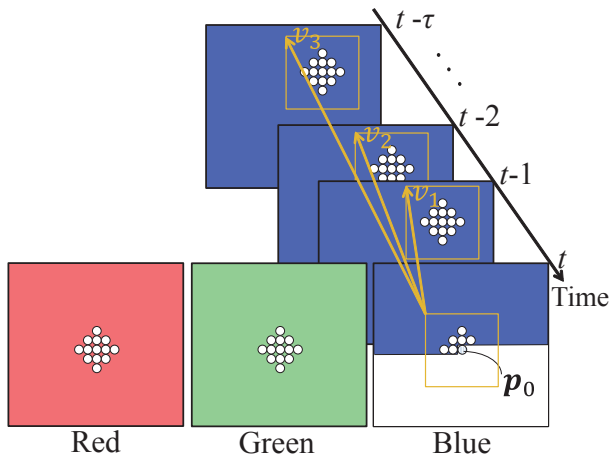


図 1 B 信号符号化時の予測手法



図 2 参照画素の配置

に符号化される R 信号に対しては、モノクローム動画像を対象とした方式 [3] と同様な予測手順を適用する。具体的には、フレーム毎に複数の線形予測器を用意し、可変サイズのブロック単位で適応選択することで、画像の局所的な性質に応じた予測を実現している。また、各線形予測器の参照画素群は、当該フレームの符号化済み近傍領域だけでなく、MC が適用された過去 3 枚のフレーム上にも配置されており、これらの線形結合により予測値を算出することで、時間方向と空間方向の相関に起因した冗長度を同時に削減可能である。

一方、G、B 信号の符号化に際しては、上記の線形予測器を、同一時刻の符号化済み色信号も参照するように拡張している [4]。例えば、B 信号符号化時の参照画素群は、図 1 に白丸で示したように符号化対象信号の符号化済み画素だけでなく、符号化済みの他信号も含めた最大 6 種類のフレーム上に配置されることとなる。RGB 4:4:4 形式のカラー動画像の色信号間の冗長度削減策としては、これまでに適応的な色座標変換 [5] や色信号間のアフィン予測 [6] などを、他の予測処理とは独立に適用する手法が検討されている。これに対し提案方式では、時間方向、空間方向、色信号間の相関を同一の予測処理の中で利用することが可能であり、予測係数を適切に決定することでより効果的に冗長度を削減できるようになると考えられる。ここで、時刻 t における符号化対象色信号を $s_t() \in \{R_t(), G_t(), B_t()\}$ と表記すると、 m 番目の予測器 ($m \in \{1, 2, \dots, M\}$) が選択

されたブロックに属する画素 p_0 の予測値は次式で表せる。

$$\begin{aligned} \hat{s}_t(p_0) = & \sum_{k=1}^{K_0} a_m(k) \cdot s_t(p_k) \\ & + \sum_{i=1}^3 \sum_{k=1}^{K_i} b_m(i, k) \cdot s_{t-\tau_i}(q_{k-1} + v_i) \\ & + \sum_{k=1}^{K_4} c_m(1, k) \cdot s'_t(q_{k-1}) + \sum_{k=1}^{K_5} c_m(2, k) \cdot s''_t(q_{k-1}) \end{aligned} \quad (1)$$

但し、 $a_m()$ 、 $b_m()$ はそれぞれ現在および過去の符号化対象色信号、 $c_m()$ は $s_t()$ と同一時刻の符号化済み色信号に配置された参照画素群に対する予測の重み (予測係数) である。また、 p_k 、 q_k は、各参照画素群の相対位置を表すベクトルであり、それぞれ図 2 のようにらせん状に配置されている。なお、(1) 式における色信号 ($s_t()$ 、 $s'_t()$ 、 $s''_t()$) の組み合わせは、 $(R_t(), -, -)$ 、 $(G_t(), R_t(), -)$ 、 $(B_t(), G_t(), R_t())$ のいずれかであり、参照すべき色信号が存在しない場合 (“-” と表記) は右辺第 3 項または第 4 項以降が不要となる。さらに、符号化区間の先頭フレームなど、過去の符号化済みフレームが参照できない場合は、右辺第 2 項の一部が省略される。

2.2 可変ブロックサイズの導入

上で述べたように、(1) 式の右辺第 2 項は過去の符号化済みフレームを用いた MC 予測に対応している。文献 [3] では、H.264/AVC の複数参照フレームを用いた双予測方式 [8] と同様に、最大 2 枚の過去フレームを同時に参照していた。これに対し、より多くのフレームを同時参照することで、動ベクトルや参照フレームの選択情報は増加するものの、予測精度を改善可能であるとの指摘もある [7]。そこで本稿では、同時に参照するフレームを 3 枚に拡張すると共に、うち 2 枚については参照フレームの時刻を直近となる $\tau_1 = 1, \tau_2 = 2$ に固定し、残る 1 枚のみ過去の 6 枚から適応選択することとした ($\tau_3 \in \{1, 2, \dots, 6\}$)。また、これらの参照フレームの MC に必要な 3 本の動ベクトル v_1, v_2, v_3 (各々整数画素精度) は、[3] と同様に 3 段階のサイズ (32×32 画素、 16×16 画素、 8×8 画素) に四分木分割されたブロック (以後 MC ブロックと呼ぶ) 単位で検出する。

一方、予測器の切り替えを行うブロック (以後予測ブロックと呼ぶ) のサイズについては、これまで各予測ブロックが MC ブロックに内包されることを保証するため一律に 8×8 画素としていた [3]。このため、オブジェクト境界などテクスチャの変化が激しい領域では予測精度の低下を招いていた。そこで本稿では、予測器の適応選択の手順を見直し、予測ブロックについても可変サイズの分割を可能とする手法の導入を図る。具体的には、 16×16 画素のブロックを最大サイズとし、これを 4×4 画素になるまで再帰的

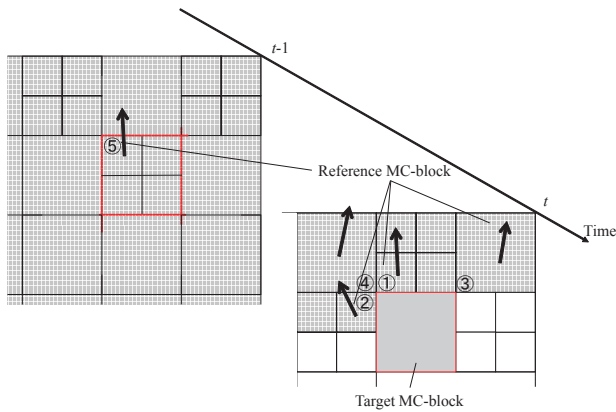


図 3 R 信号の動ベクトル符号化時の参照動ベクトル

に 4 分木分割しながら、それぞれのブロックに適切な予測器の選択を行う。この際に、予測誤差の符号量と付加情報の符号量の総和の値をコスト関数とし、このコスト関数が最小となるような分割パターンを探索している [9]。

3. 符号化手順

3.1 動ベクトルの符号化

提案方式では、各色信号の MC ブロックごとに 3 本の動ベクトル $v_i (i = 1, 2, 3)$ を付加情報として伝送する必要がある。H.264/AVC などの国際標準方式では、符号化済み近傍ブロックの参照動ベクトルからメディアン予測などによって予測動ベクトルを求め、実際の動ベクトルとの差分ベクトルを符号化対象とするのが一般的である。これは、差分動ベクトルの水平・垂直成分の頻度分布が 0 付近にピークを持つ単峰性の確率モデルに従うことを仮定し、効率的なビット割当てを行っているとして解釈できる。しかし、動きの異なるオブジェクトの境界部などでは、符号化対象動ベクトルがそれぞれのオブジェクトの動きに一致する確率が有意に高くなると考えられ、単峰性の確率モデルではミスマッチを生じてしまう可能性がある。そこで提案方式では、全ての参照動ベクトルを反映した多峰性の確率モデルに基づいて、符号化対象動ベクトルの水平・垂直成分を直接符号化する手法 [10] を導入し、付加情報の削減を図る。なお以下では簡単のため動ベクトル v_i の水平成分 v_x の符号化について述べるが、垂直成分 v_y の符号化方法も同様である。

まず、R 信号の動ベクトルに関しては、図 3 に示すように当該フレームの近傍 MC ブロック ①～④および前フレームの左上隅に位置する MC ブロック ⑤の符号化済み動ベクトルを参照動ベクトルとする。但し、符号化対象動ベクトルが v_3 の場合は、各動ベクトルの時間的な距離 τ_3 が一致しない場合があるため、参照動ベクトルの長さに関して正規化処理 [11] を施しておく。ここで、各参照動ベクトルの正規化後の水平成分を $\mathbf{V} = \{v_x(1), v_x(2), \dots, v_x(5)\}$ とおき、符号化対象となる v_x の条件付き確率分布を次式で定

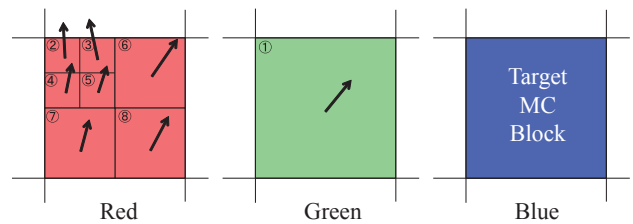


図 4 B 信号の動ベクトル符号化時の参照動ベクトル

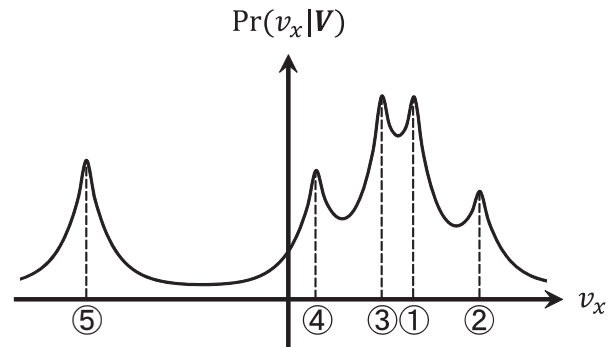


図 5 多峰性確率モデルの例

義する。

$$\Pr(v_x|\mathbf{V}) \propto \sum_{j=1}^{|\mathbf{V}|} w_j \cdot g(v_x - v_x(j)) \quad (2)$$

但し、 w_j はそれぞれの参照動ベクトルを仮定してブロックマッチングによる予測画像を生成した際の、ブロック境界の整合性に基づいて算出された重み係数である [10]。また、 $g()$ は一般化ガウス関数を表し、その標準偏差および形状パラメータはフレーム毎に適切な値を決定している。したがって、 $\Pr(v_x|\mathbf{V})$ は図 5 のように最大 5 箇所ピークを持つ多峰性の確率分布となり、これを多値算術符号器の確率モデルとして使用することで、より適切な符号割当てが可能になると期待できる。

次に、G, B 信号の動ベクトル符号化の際には、各色信号の動き情報は類似していると考えられるため、当該ブロックを符号化済み色信号上に重ね、その内部に含まれる MC ブロック全て (図 5 の例では ①～⑧) から参照動ベクトル $\mathbf{V} = \{v_x(1), v_x(2), \dots\}$ を収集する。このとき、(2) 式における重み係数 w_j は当該ブロック内部に関して、参照動ベクトルが属する MC ブロックの面積が占める割合とする。

3.2 予測誤差の符号化

2.1 で述べた予測処理の結果得られた予測誤差信号に対しては、画素ごとのコンテキストモデリングに基づいた適応的な多値算術符号化が適用される。コンテキストモデリングとは、周囲の画素の状態 (コンテキスト) から、符号化対象画素の予測誤差の確率密度関数の推定を行う手法である [12]。提案方式では、空間方向および時間方向の近傍画素群の予測誤差信号の絶対値和を表す特徴量 $U(\mathbf{p}_0)$ を、し

きい値 $Th_m(1), Th_m(2), \dots, Th_m(15)$ を用いて 16 レベルに量子化した値によってコンテキスト $n \in \{1, 2, \dots, 16\}$ を定義している。

$$U(\mathbf{p}_0) = \sum_{k=1}^6 |s_t(\mathbf{p}_k) - \hat{s}_t(\mathbf{p}_k)| + \sum_{i=1}^2 \sum_{k=1}^5 |s_{t-i}(\mathbf{q}_{k-1} + \mathbf{v}_i) - \hat{s}_{t-i}(\mathbf{q}_{k-1} + \mathbf{v}_i)| \quad (3)$$

また、 n 番目のコンテキストに分類された画素における予測誤差信号の確率密度関数を、次式の一般化ガウス関数でモデル化する。

$$P(e|n) = \frac{c_n \eta(c_n, \sigma_n)}{2\Gamma(1/c_n)} \cdot \exp\left\{-|\eta(c_n, \sigma_n) \cdot e|^{c_n}\right\},$$

$$\eta(c_n, \sigma_n) = \frac{1}{\sigma_n} \sqrt{\frac{\Gamma(3/c_n)}{\Gamma(1/c_n)}}, \quad (4)$$

但し、 $\Gamma(\cdot)$ はガンマ関数、 σ_n は誤差 e の標準偏差、 c_n は一般化ガウス関数の分布形を制御するパラメータ (形状パラメータ) である。提案方式では、16 通りの σ_n の値を先に与えておき、同一の予測器が割り当てられる領域ごとにコンテキストモデリングのしきい値 $Th_m(1), Th_m(2), \dots, Th_m(15)$ を最適化することで、 $P(e|n)$ を実際の分布形に近づけている。また、形状パラメータ c_n の値についても各フレームの色信号ごとに適切な値を設定している。ここで、符号化対象信号 $s_t(\mathbf{p}_0)$ が 8 ビットの整数値であることに注意すると、予測値 $\hat{s}_t(\mathbf{p}_0)$ が既知の場合の予測誤差 $e = s_t(\mathbf{p}_0) - \hat{s}_t(\mathbf{p}_0)$ の取り得る値は 256 通りに制約される。それぞれの生起確率 $\Pr(e|\hat{s}_t(\mathbf{p}_0), n)$ を (4) 式に基づいて計算しておくことで、予測値とコンテキストが与えられた際の実際の予測誤差 e の符号量を次式で見積もることができる [12]。

$$L(e|\hat{s}_t(\mathbf{p}_0), n) = -\log_2 \Pr(e|\hat{s}_t(\mathbf{p}_0), n) \quad (5)$$

提案方式では、上述したコンテキストモデリングのしきい値や一般化ガウス分布の形状パラメータの他にも、以下のパラメータを付加情報として符号化する必要があるが、これらは全て (5) 式の和が最小となるように各フレームの色信号ごとに繰り返し最適化を図っている。

- 各予測器を構成する予測係数 $a_m(), b_m(), c_m()$
- MC ブロックの分割状況
- 各 MC ブロックの動ベクトル $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ および参照フレーム番号 τ_3
- 予測ブロックの分割状況
- 各予測ブロックで選択された予測器の番号 m

さらに、各予測器の予測回数に関するパラメータ K_0, K_1, \dots, K_5 および各フレームの色信号ごとに用意される予測器の数 M についても、符号化開始時に十分大きな値を設定しておき、上記パラメータの最適化手順の中で

表 1 符号化レート (bits/pel)

Image	提案方式	固定ブロック	H.264
Truck train	8.472	8.593	10.408
Red leaves	10.843	10.859	12.957
Fountain	9.223	9.236	13.419
Studio concert	11.738	11.790	13.968
Woman at harbor	8.326	8.332	9.654
Colorful world	11.031	11.098	13.381
Average	9.939	9.972	12.298

表 2 動画 Truck train の符号量の内訳 (bits/frame)

	可変ブロック	固定ブロック
予測誤差	812611.5	823589.8
予測係数	13658.6	14095.4
予測器の選択情報	18745.3	13915.5
動き情報	5670.9	6235.9
その他	2426.0	2473.32
Total	855852.5	863099.9

徐々に削減することで適切な値を得る手法 [13] を導入している。

4. 特性評価

提案方式の性能を評価するため、ITE/ARIB ハイビジョン・システム評価用標準動画図第 2 版から抜粋し、 352×288 画素に変換した RGB 4:4:4 カラー画像 (各色 8bit 精度、25 フレーム分) を対象として符号化実験を実施した。表 1 にその結果を示す。但し、「固定ブロック」は、提案方式において予測ブロックのサイズを 8×8 画素に固定した場合の性能を表している。また、「H.264」は H.264/AVC の High 4:4:4 プロファイルに基づいた可逆符号化方式 (IPBBB, 参照フレーム 6 枚, CABAC 符号化)[14] を用いて、R, G, B 信号を独立に符号化した結果を表している。これより、予測ブロックサイズを可変とした提案方式は、固定ブロックサイズの場合と比較して全ての動画像において符号化レートを改善している様子を確認できる。

次に、シーケンス Truck train における、符号量の内訳を表 2 に、提案方式における予測ブロックの分割状況を図 6 に示す。表 2 より、予測器の選択情報に要する符号量が増大しているが、それ以上に予測誤差の符号量の減少幅が大きいことがわかる。これは、図 6 のように、テクスチャが複雑な領域やオブジェクトの境界部が比較的小さなブロックに分割されたことにより、予測精度が向上したためであると考えられる。

5. まとめ

本稿では、RGB 4:4:4 フォーマットのカラー動画像を対象とした高効率な可逆符号化方式の提案を行った。提案方式では、時間方向、空間方向、色信号間の相関を同時に利用可能な予測器を、可変サイズのブロック単位で切り替え

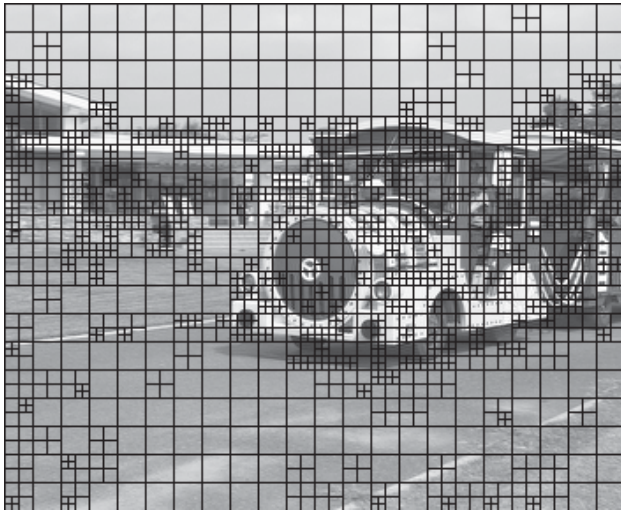


図 6 予測ブロックの分割状況 (Truck train R 信号 7 フレーム目)

る事で画像の局所的な性質に追従した高精度な予測を実現している。また、同時に3枚のフレームを用いるMC予測の採用に伴い、増加した動ベクトルの符号量を抑えるため、多峰性確率モデルを用いた算術符号化手法を導入した。符号化実験の結果、予測ブロックサイズを可変とすることでこれを固定とした場合に比較して平均で0.033bits/pel符号化レートが改善される事を確認した。また、動画像符号化国際標準方式であるH.264/AVCに基づいた可逆符号化方式と比較して、13.7%–31.2%低い符号化レートを達成することが示された。

参考文献

- [1] Wiegand, T., Sullivan, G.J., Bjøntegaard, G. and Luthra, A.: Overview of the H. 264/AVC Video Coding Standard, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp.560–576 (2003).
- [2] Sullivan, G.J., Ohm, J., Han, W.J. and Wiegand, T.: Overview of the High Efficiency Video Coding (HEVC) Standard, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 12, No. 12, pp.1649–1668 (2012).
- [3] Maeda, H., Minezawa, A., Matsuda, I. and Itoh, S.: Lossless Video Coding Using Multi-frame MC and 3D Bi-prediction Optimized for Each Frame, *Proc. of European Signal Processing Conference (EUSIPCO 2006)*, Thu. 3.2 (2006).
- [4] Matsuda, I., Kaneko, T., Minezawa, A. and Itoh, S.: Lossless Coding of Color Images Using Block-Adaptive Inter-Color Prediction, *Proc. of IEEE International Conference on Image Processing (ICIP 2007)*, Vol. 2, pp.329–332 (2007).
- [5] Marpe, D., Kirchhoffer, H., George, V., Kauff, P. and Wiegand, T.: Macroblock-adaptive Residual Color Space Transforms for 4:4:4 Video Coding, *Proc. of IEEE International Conference on Image Processing (ICIP 2006)*, pp.3157–3160 (2006).
- [6] 加藤晴久, 米山暁夫, 内藤 整, 酒澤茂之, 松本 修一: H.264のIntra予測残差に対する適応的チャンネル間予測, 映像情報メディア学会誌, Vol. 64, No. 11, pp.1711–1717 (2010).
- [7] Flierl, M., Wiegand, T. and Girod, B.: Generalized B Pictures and the Draft H.264/AVC Video-compression

- Standard, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 12, No. 11 (2002).
- [8] Flierl, M. and Girod, B.: Rate-constrained Multi-hypothesis Motion-compensated Prediction for Video Coding, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp.587–597 (2003).
- [9] Matsuda, I., Ozaki, N., Umezumi, Y. and Itoh, S.: Lossless Coding Using Variable Block-size Adaptive Prediction Optimized for Each Image, *Proc. of European Signal Processing Conference (EUSIPCO 2005)*, WedAmPO3 (2005).
- [10] 伊藤祐也, 青森 久, 松田 一朗, 伊東 晋: 多峰性確率モデルを用いた動ベクトルの算術符号化に関する基礎検討, 電子情報通信学会技術研究報告, Vol. 111, No. 284, pp.19–24 (2011).
- [11] Kato, S. and Boom, C. S.: *Motion Vector Prediction for Multiple Reference Frame Video Coding Using Temporal Motion Vector Normalization*, 第19回画像符号化シンポジウム (PCSJ 2004), P-2.18 (2004).
- [12] Matsuda, I., Shiodera, T. and Itoh, S.: Lossless Video Coding Using Variable Block-size MC and 3D Prediction Optimized for Each Frame, *Proc. of European Signal Processing Conference (EUSIPCO 2004)*, pp/1967–1970 (2004).
- [13] 小崎 南羽, 松田 一朗, 森岡 一幸, 伊東 晋: ブロック適応予測を用いた可逆符号化のためのパラメータ自動設定, 電子情報通信学会技術研究報告, Vol. 105, No. 432, pp. 7–12(2005).
- [14] Sullivan, G.J., Haoping Yu, Sekiguchi, S.-I., Huifang Sun, Wedi, T., Wittmann, S., Lee, Y.L., Segall, A. and Suzuki, T.: New Standardized Extensions of MPEG4-AVC/H.264 for Professional-Quality Video Applications, *Proc. of IEEE International Conference on Image Processing (ICIP 2007)*, Vol. 1, pp.I-13–I-16 (2007).