

## 映像符号化のための 動き・エッジ検出型時空間フィルタリングの検討

比企 豊<sup>†</sup> 村松 正吾<sup>†</sup> 菊池 久和<sup>†</sup>

<sup>†</sup>新潟大学工学部電気電子工学科 〒950-2181 新潟市西区五十嵐2の町8050番地

E-mail: [†mhiki@telecom0.eng.niigata-u.ac.jp](mailto:†mhiki@telecom0.eng.niigata-u.ac.jp)

**あらまし** 本報告では、低消費電力が要求される映像符号化のための新たなツールとして、動き・エッジ検出型予測処理を用いた非分離時空間フィルタリングについて検討する。本手法は過去に筆者らが提案してきた非分離動き補償時空間フィルタリング (MCSTF) における動き補償予測処理を動き・エッジ検出型予測処理に置き換えたものである。提案手法は予測処理に動き補償を用いないため、演算コストを抑えることができるという利点を持つ。また、リフティングで構成されるために、完全再構成を保証する。JPEG2000 を用いた初期実験モデルにより復号後の PSNR を比較し、本手法の有効性を確認する。

**キーワード** 動き補償時間フィルタリング, 映像符号化, エッジ適応補間

## Spatio-temporal Filtering with Motion/Edge-based Prediction for Video Coding

Minoru HIKI<sup>†</sup>, Shogo MURAMATSU<sup>†</sup>, and Hisakazu KIKUCHI<sup>†</sup>

<sup>†</sup> Faculty of Engineering, Niigata University Ikarashi 2-no-cho, Nishi-ku, Niigata, 950-2181 Japan

E-mail: [†mhiki@telecom0.eng.niigata-u.ac.jp](mailto:†mhiki@telecom0.eng.niigata-u.ac.jp)

**Abstract** This work studies spatio-temporal filtering with motion/edge-based prediction for video coding, which is an alternative technique of non-separable motion-compensated spatio-temporal filtering (MCSTF). The proposed technique replaces motion compensation of MCSTF with edge-based prediction, and has an advantage in computational efficiency without motion-compensation. The proposed technique is based on the lifting scheme so that perfect reconstruction property is guaranteed. Some experimental results with JPEG2000 show the significance of our proposed technique.

**Key words** Motion-compensated temporal filtering, Video coding, Adaptive edge interpolation

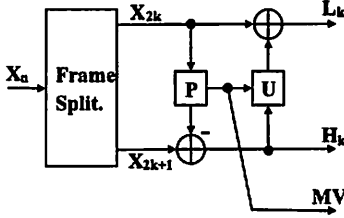
### 1. はじめに

ネットワークの普及、ワイヤレス通信技術、映像圧縮技術の発展により、ウェブカメラを利用したワイヤレス映像監視システムが実用化されている。また、自然監視など、電力の確保が難しい環境化にある映像監視システムにおいては、特に低コスト・低消費電力の処理が要求される [1]。しかしながら、現在の MPEG-4 や H.264/MPEG-4 AVC などの主要な映像符号化は動き補償を用いているため、ワイヤレス映像監視システムにおける映像圧縮に最適であるとはいえない。次世代映像符号化の要素技術として注目されている動き補償時間フィルタリング (MCTF) においても、動き補償を用いたリフティング処理を行っているため、従来の動き補償と比較してさらに演算量は増加している [2-4]。そこで筆者らは、低消費電力が要求される

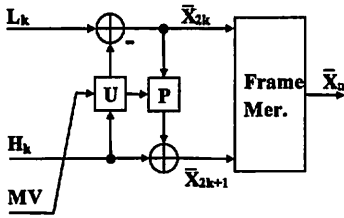
システムのための映像符号化の要素技術として、新たにエッジ検出型非分離時空間フィルタリング (EDSTF) を提案した [5]。EDSTF は過去に筆者らが提案してきた、MCTF の時空間拡張である非分離動き補償時空間フィルタリング (MCSTF) [6-8] における動き補償予測処理部分にエッジ検出型予測処理を適用したものである。

提案手法は動き補償を用いないため、演算コストを抑えることができるという利点を持つ。また、リフティングで構成されるために、完全再構成を保証する。しかしながら、従来の EDSTF では空間補間を行うため、静止部分において性能が低下してしまうという問題があった。そこで本報告では、EDSTF に動き検出を加えた動き・エッジ検出型予測処理を提案する。JPEG2000 と組み合わせたシミュレーションモデルにより、その画質の評価を行い、提案手法の有効性に関する検討を行う。

## 2. 動き補償時間フィルタリング



(a) 順変換



(b) 逆変換

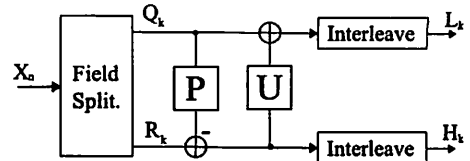
図1 動き補償時間フィルタリング

図1に動き補償時間フィルタリング(MCTF)の基本構成を示す。MCTFは、リフティング構成で実装され、従来のクロスストループ型の動き補償と比較してドリフトが存在しないなどの利点を持つ。リフティングは一般的に、フレーム分割処理(Frame split.)、予測部(P)、及び更新部(U)の3つのステップから構成される。まず、入力される映像信号(フレーム列)は、時間方向に偶数フレーム、奇数フレームに分割される。次に、予測部により、隣接する偶数フレームから奇数フレームとの予測処理をし、ハーフレートの高域フレームを生成する。そのあと、更新部において、その高域フレームからハーフレートの低域フレームが得られる。動き補償の適用により、この予測部での予測誤差を低減し、符号化効率を上げることが可能である。

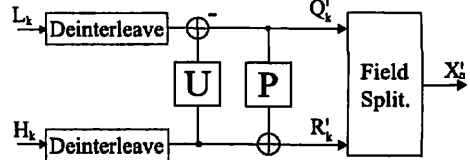
MCTFは時間スケラビリティ、SNRスケラビリティなどの点で有益な符号化構造である。H.261/263、MPEG-1/2/4などに代表される映像符号化で用いられていた再帰的な予測処理(Closed-loop)は、ドリフトの発生が問題となっていたが、MCTFは非再帰的な予測構造(Open-loop)であるため、ドリフトが発生しない。しかしながら、MCTFは更新処理なしで構成した場合、フレームごとPSNRの揺らぎや符号化効率の減少が問題となる。また、更新処理を加えて構成した場合、処理遅延・演算コストの増加などの問題を抱えている。

## 3. 非分離時空間フィルタリング

図2に非分離時空間フィルタリングの構成を示す。非分離時空間フィルタリングはフィールド分割処理(Field Split.)、予測部(P)、更新部(U)、インタリーブ処理から構成される。フィールド分割により生成された2つのフィールド列は、予測部に入力され、予測誤差画像を生成する。本報告の実験では、単純化のた



(a) VT-Quincunx 格子 (ラインベース)



(b) FCO 格子 (ポイントベース)

図2 フィールド分割処理

めに更新部を省略した。このフィールド列に対するリフティング構成を行った後、インタリーブ処理をする。本手法はMCTFや過去に提案してきたMCSTFと異なり、動き補償を用いずに予測処理を行うため、動きベクトルを伝送する必要がなく、演算コストが少ないという利点がある。

### 3.1 フィールド分割処理

図3(a)にVT-Quincunx格子(ラインベース)、3(b)にFCO格子(ポイントベース)のフィールド分割処理をそれぞれ示す。フィールド分割処理では、入力フレーム列 $X_n$ は、二つのフィールド列 $Q_k, R_k$ に分割される。図中の $n$ はフルレートのフレーム番号、 $k$ はハーフレートのフレーム番号をそれぞれ示している。 $P_{Tn}, P_{Bn}$ はそれぞれ $P_n$ のトップフィールドとボトムフィールドを示し、 $Q_k = P_{T2k}, P_{B(2k+1)}$ 、 $R_k = P_{T(2k-1)}, P_{Bk}$ と関係付けられる。フィールド分割は従来のフレーム分割と異なり、時間方向遅く処理に加え、空間方向の相関を用いた予測処理も可能となる。

### 3.2 エッジ検出型予測処理

#### 3.2.1 VT-Quincunx 格子 (ラインベース)

図4(a)にラインベースエッジ検出予測処理を示す。図??(b)はトップフィールドからボトムフィールドの画素を予測する場合である。始めに、垂直方向、エッジが左斜め方向、右斜め方向のいずれの方向であるかを推定する。垂直方向のインジケータ $\alpha$ を

$$\alpha = |T_{21} - T_{23}|, \quad (1)$$

左斜めのインジケータ $\beta$ を

$$\beta = |T_{11} - T_{33}|, \quad (2)$$

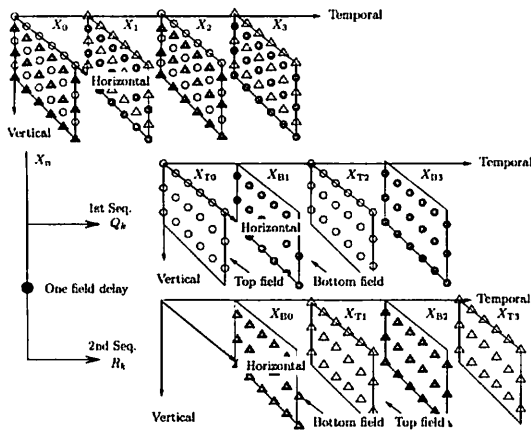
右斜めのインジケータ $\gamma$ を

$$\gamma = |T_{31} - T_{13}|, \quad (3)$$

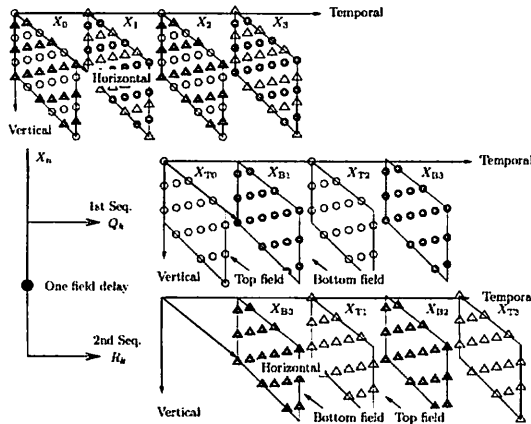
と定義する。

a)  $\alpha < \beta, \gamma$  のとき

垂直方向の変化が小さいので、垂直にエッジが走っていると



(a) VT-Quincunx 格子 (ラインベース)



(b) FCO 格子 (ポイントベース)

図3 フィールド分割処理

推定される。したがって、エッジに沿ってボトムフィールドの画素を予測するために、次式を用いて垂直方向から  $B_{22}$  を予測する

$$B_{22} = \frac{T_{21} + T_{23}}{2}, \quad (4)$$

b)  $\beta < \alpha, \gamma$  のとき

同様に、左斜め方向の変化が小さいので、右斜めにエッジが走っていると推定される。したがって、エッジに沿ってボトムフィールドの画素を予測するために、次式を用いて右斜め方向から  $B_{22}$  を予測する。

$$B_{22} = \frac{T_{11} + T_{33}}{2}, \quad (5)$$

c)  $\gamma < \alpha, \beta$  のとき

右斜め方向の変化が小さいので、右斜めにエッジが走っていると推定される。したがって、エッジに沿ってボトムフィールドの画素を予測するために、次式を用いて右斜め方向から  $B_{22}$  を

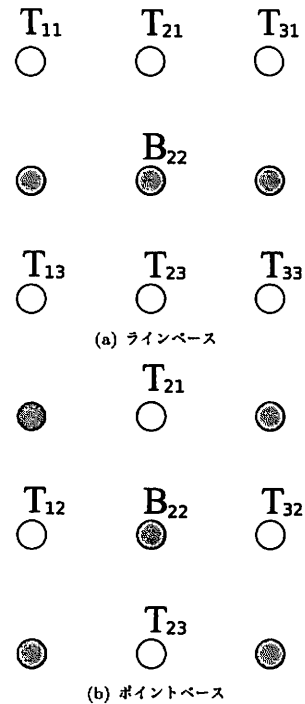


図4 エッジ検出型予測処理

予測する。

$$B_{22} = \frac{T_{31} + T_{13}}{2}, \quad (6)$$

### 3.2.2 FCO 格子 (ポイントベース)

ラインベースエッジ検出予測処理と同様に、図4(b)にポイントベースエッジ検出遅く処理を示す。始めに、エッジが垂直方向、水平方向のいずれか方向であることを推定する。垂直方向のインジケータ  $\alpha$  を

$$\alpha = |T_{21} - T_{23}|, \quad (7)$$

水平方向のインジケータ  $\beta$  を

$$\beta = |T_{12} - T_{32}|, \quad (8)$$

と定義する。

a)  $\alpha < \beta$  のとき

垂直方向の変化が小さいので、垂直にエッジが走っていると推定される。したがって、エッジに沿ってボトムフィールドの画素を予測するために、次式を用いて垂直方向から  $B_{22}$  を予測する。

$$B_{22} = \frac{T_{21} + T_{23}}{2}, \quad (9)$$

b)  $\alpha > \beta$  のとき

$$B_{22} = \frac{T_{12} + T_{32}}{2}, \quad (10)$$

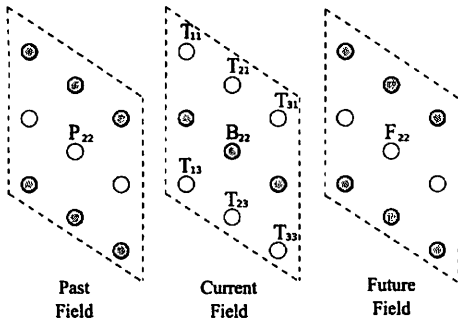


図 5 動き・エッジ検出型予測処理

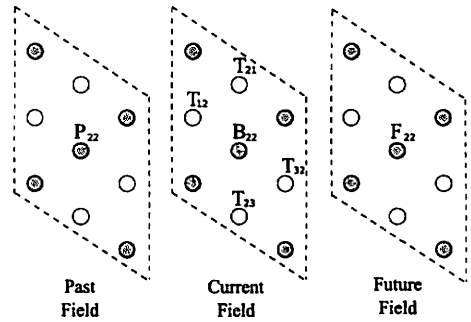


図 6 動き・エッジ検出型予測処理

### 3.3 動き・エッジ検出型予測処理

エッジ検出型予測処理では予測を空間方向の予測処理を行う。しかし、静止部分においては時間方向の予測処理の方が予測誤差を低減させることができる。動き・エッジ検出型予測処理は、エッジ検出に時間方向の勾配も加えて処理を行う。

#### 3.3.1 ラインベース

図 5 にラインベースの動き・エッジ検出型予測処理を示す。始めに、エッジ検出型予測処理と同様に、エッジが垂直方向、水平方向のいずれに走っているか、もしくは動きがあるかどうかを推定する。垂直方向のインジケータ  $\alpha$  は式 (1)、左斜め方向のインジケータ  $\beta$  は式 (2)、右斜め方向のインジケータ  $\gamma$  は式 (3) をそれぞれ使用する。また、動きを検出するためのインジケータ  $\delta$  を

$$\gamma = |P_{22} - F_{22}|, \quad (11)$$

と定義する。

a)  $\alpha < \beta, \gamma, \delta$  のとき

エッジ検出型予測処理と同様に、式 (4) を用いて  $B_{22}$  を予測する。

b)  $\beta < \alpha, \gamma, \delta$  のとき

同様に、式 (5) を用いて  $B_{22}$  を予測する。

c)  $\gamma < \alpha, \beta, \delta$  のとき

同様に、式 (6) を用いて  $B_{22}$  を予測する。

d)  $\delta < \alpha, \beta, \gamma$  のとき

時間方向の変化が少ないので、動きがないと推定される。したがって、ボトムフィールドの画素を予測するために、次式を用いて前後フレームの同位置にある画素から  $B_{22}$  を予測する。

$$B_{22} = \frac{P_{22} + F_{22}}{2}, \quad (12)$$

#### 3.3.2 ポイントベース

図 6 にポイントベースの動き・エッジ検出型予測処理を示す。始めに、エッジ検出型予測処理と同様に、エッジが垂直方向、水平方向のいずれに走っているか、もしくは動きがあるかどうかを推定する。垂直方向のインジケータ  $\alpha$ 、水平方向のインジケータ  $\beta$  は式 (7)、(8) をそれぞれ使用する。また、動きを検出するためのインジケータ  $\gamma$  を

$$\gamma = |P_{22} - F_{22}|, \quad (13)$$

と定義する。

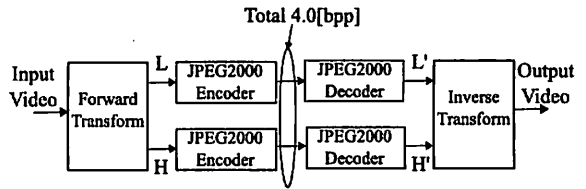


図 7 シミュレーションモデル

a)  $\alpha < \beta, \gamma$  のとき

エッジ検出型予測処理と同様に、式 (9) を用いて  $B_{22}$  を予測する。

b)  $\beta < \alpha, \gamma$  のとき

同様に、式 (10) を用いて  $B_{22}$  を予測する。

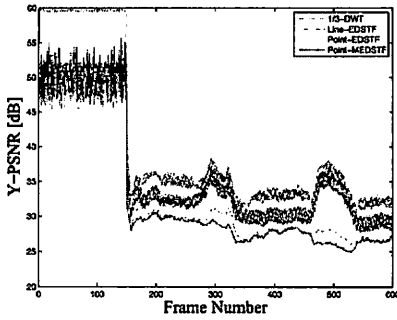
c)  $\gamma < \alpha, \beta$  のとき

時間方向の変化が少ないので、動きがないと推定される。したがって、ボトムフィールドの画素を予測するために、次式を用いて前後フレームの同位置にある画素から  $B_{22}$  を予測する。

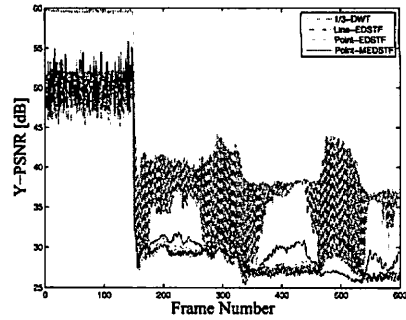
$$B_{22} = \frac{P_{22} + F_{22}}{2}, \quad (14)$$

## 4. 性能評価

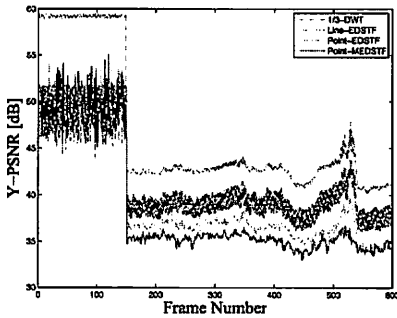
提案手法の有効性を検討するために、符号化性能を評価する。図 7 にシミュレーションモデルを示す。本報告では時間方向 1/3-DWT、ラインベース、ポイントベースのエッジ検出型予測処理 (Line-EDSTF, Point-EDSTF)、動き・エッジ検出型予測処理 (Line-MEDSTF, Point-MEDSTF) による時間方向分割を 1 段行う。その後、高域成分に対してそれぞれビットレートを指定して JPEG2000 を用いて符号化、復号を行い再構成をして元の画像との PSNR を評価する。使用する動画は、ITE 標準動画 (SIF *Whale Show*, *Crowded Crosswalk*, 352 x 240, 30 frames/sec., 8-bit グレイスケール, 600 フレーム) とする。*Whale Show* はパン・ズームを含みカメラワークが激しく、*Crowded Crosswalk* は固定カメラで撮影されたものである。動き補償に関しては、全探索アルゴリズム、ブロックサイズは 16 x 16 サイズとした。また、JPEG2000 符号化、復号には JJ2000 を用いた [9]。図 8 にフレームごとの PSNR を示す。図 8 (a), (c) は時間方向低域に 2.0 [bpp]、高域に 2.0 [bpp] を割り当てたとき。図 8 (b), (d) は時間方向低域に 3.0 [bpp]、高



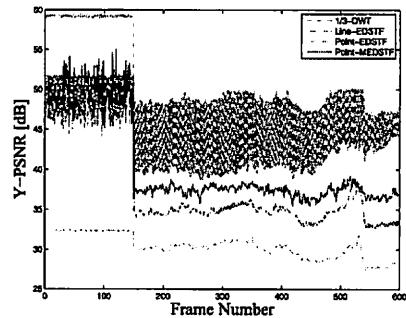
(a) Whale Show@2.0, 2.0



(b) Whale Show@3.0, 1.0



(c) Crowded Crosswalk@2.0, 2.0



(d) Crowded Crosswalk@3.0, 1.0

図8 フレーム毎のPSNR

域に 1.0 [bpp] 割り当てたときである。

図 8 (a), (b) の動きの激しい部分において、高域にビットを多く割り当てたときにラインベースの提案手法の性能が 1/3-MCTF に迫る性能が得られていることがわかる。Crowded Crosswalk においては、低域にビットを多く割り当てたときエッジ検出型予測処理に時間方向補間を加えた予測処理の性能があがっていることがわかる。したがって、動きが少ない映像に対しては時間方向の補間を加えた予測処理が適していると言える。すべての結果において、時間方向 1/3-DWTF よりも性能が低下しているが、これはインタリーブ画像に対して JPEG2000 で符号化をしているためと思われる。変換後のレート制御も今後の検討課題である。

## 5. おわりに

映像符号化のための動き・エッジ検出型予測処理を用いた非分離時空間フィルタリングを提案した。JPEG2000 を用いたシミュレーションモデルを用いて、フレームごとの PSNR を比較し、本手法の検討を行った。今後の課題は、予測方法の検討、歪レート制御、エントロピー符号化の導入と性能評価である。

## 文 献

- [1] 岩橋政宏, S. Udomsiri, 今井裕二, “機能的階層符号化と河川監視への応用,” 映像情報メディア学会メディア工学研究会, vol. 30, no. 68, pp.77-80, Dec. 2006.
- [2] J.-R. Ohm, “Advances in scalable video coding,” Proceedings of the IEEE, vol. 93, no. 1, pp. 42-56, Jan. 2005.
- [3] A. Secker and D. Taubman, “Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression,” IEEE Trans. Image Processing, vol. 3, no. 5, pp. 559-571, Sept. 1994.
- [4] R. Schafer, H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, “MCTF and scalable extension of H.264/AVC and its application to video transmission, storage, and surveillance,” in Proc. SPIE, Visual Commun. Image Process., pp. 343-354, Beijing, China, July 2005.
- [5] 比企豊, 村松正吾, 菊池久和, “映像符号化のためのエッジ検出型予測処理を用いた時空間フィルタリングの検討,” 電子情報通信学会信号処理研究会, pp. 37-41, Apr. 2007.
- [6] M. Hiki, T. Ishida, S. Muramatsu, and H. Kikuchi, “Suppression of PSNR fluctuation in motion-compensated temporal 1/3-transform through non-separable sub-sampling,” Proc. IEEE ICIP 2006, pp. 2153-2156, Atlanta, G.A., U.S.A., Oct. 2006.
- [7] 比企豊, 石田拓磨, 村松正吾, 菊池久和, “フィールド・スケラブル映像符号化のための非分離 MCTF の基礎検討,” 第 20 回映像符号化シンポジウム, pp.71-72, Nov. 2005.
- [8] 村松正吾, 比企豊, 石田拓磨, 菊池久和, “動き補償時空間フィルタリングにおける標準化格子の検討,” 映像情報メディア学会メディア工学研究会, pp. 5-8, Dec. 2005.
- [9] Cannon, EPFL and Ericsson, <http://ij2000.epfl.ch>, URL.