

# 画素毎の動き推定に基づく時間外挿フレームを用いた 複数参照フレーム動き補償予測

亀田 裕介<sup>†1,a)</sup> 岸 浩志<sup>†1</sup> 石川 知一<sup>†1</sup> 松田 一郎<sup>†1</sup> 伊東 晋<sup>†1</sup>

概要：本論文では画素毎の動き（オプティカルフロー）による時間外挿フレームを用いた複数参照フレーム動き補償予測方式を提案する。復号済みの連続2フレームから僅かな符号量で推定できる画素毎の動きを用いることで、ブロックごとの動きでは必ずしも適切ではない局所的な回転・拡大縮小・変形などの動きを補正することができる。ただし、外挿フレームは等速直線運動を仮定しており、またオクルージョンの悪影響もある。外挿フレームを参照ピクチャリストに追加して従来同様にブロックマッチングによる動き補償を行うことで、外挿フレームの欠点を補うことができる。可逆符号化での複数参照フレームの動き補償予測の性能評価の結果、外挿フレームを用いる提案手法の有効性と領域ごとの特性が確認できた。

## Multi-Frame Motion Compensation using Extrapolated Frame by Apparent Motion

YUSUKE KAMEDA<sup>†1,a)</sup> KISHI HIROYUKI<sup>†1</sup> ISHIKAWA TOMOKAZU<sup>†1</sup> MATSUDA ICHIRO<sup>†1</sup> ITOH SUSUMU<sup>†1</sup>

### 1. はじめに

映像符号化は映像の放送・通信・蓄積の基盤技術であり、2016年時点での最新の国際規格は High Efficiency Video Coding (HEVC) とよばれる H.265 である [1], [2]。映像符号化の規格において高圧縮率を実現するための基礎的技術として、フレーム間の時間的相関を利用する動き補償 (MC) 予測がある。一般的な MC では、符号化対象フレームをブロック分割し、各ブロックに対して符号化済みフレームとのブロックマッチング (BM) を行い、動ベクトル (MV) を算出する。そして、ブロック分割情報や MV などの付加情報を送信することで、符号化対象フレームを予測する。しかし、BM による MC はブロック毎の平行移動のみを補償するので、局所的な拡大縮小・回転・変形などの画素毎に異なる動きをする映像の予測に対しては必ずしも適切ではない。拡大縮小などの動きに対して高精度な

MC 予測を行うためには、実数精度の MV を画素毎に送信する必要があるが、付加情報量が増大するため非現実的である。

コンピュータビジョン分野において、画素毎の動き（オプティカルフロー）の推定法は 1980 年頃から研究されており、自動運転やロボットの視覚への応用が期待されている [3], [4]。ほとんどのオプティカルフロー推定法は変分法に基づいており、動きの見かけの境界を精度よく表現できる total variation 正則化もよく用いられている [3], [4], [5]。オプティカルフロー推定の計算については、GPGPU などを用いた画素毎の並列計算によりフル HD サイズでもリアルタイムに処理できる。そのため、オプティカルフローを用いて復号器側で復元済みのフレームである再生画像から適切に推定できれば高精度な MC を実現できると考えられる。この場合、各画素の動きを用いたフレーム内挿もしくはフレーム外挿により符号化対象フレームを近似することになる。オプティカルフロー推定法の性能評価の一つの指標はフレーム間内挿 [6] であり、フレーム内挿は映像符号化方式に応用されつつある [7]。

<sup>†1</sup> 現在、東京理科大学 (〒 278-8510 千葉県野田市山崎 2641)  
Presently with Tokyo University of Science, Yamazaki 2641,  
Noda-shi, Chiba 278-8510, Japan

<sup>a)</sup> ykameda@rs.tus.ac.jp

画素毎の見かけの動き推定に基づくフレーム外挿は、非可逆符号化においてブロックマッチングの参照フレームとの組み合わせが検討されている [8], [9]。しかし、先行研究では参照フレームが固定されており双予測に関する解析も行われていない。また、非可逆符号化のため、符号化ノイズにより外挿フレームの本質的な性能評価を実施できていなかった。そこで本論文では、可逆符号化において、画素毎の見かけの動き推定に基づく外挿フレームを用いた複数参照フレーム動き補償予測について符号量の観点から性能の評価と解析を行う。

## 2. 外挿近似のための見かけの動き推定

外挿近似フレームを生成するためには、近似対象フレームの画素毎に MV が定義される必要がある。そこで、微小時間変化において等速直線運動を仮定することで、連続する直前 2 枚の符号化済みフレームから画素毎の MV を算出する [8], [9]。本論文では画像領域  $\Omega$  で定義される画素値を  $f(\mathbf{x}, t)$  で表す。ここで、 $\mathbf{x} = (x, y)^\top \in \Omega \subset \mathbb{R}^2$  は実数画素位置であり  $t$  は時刻である。時刻  $t-2$  と  $t-1$  の符号化済みフレームについて、画素値が変わらなければ次式が成り立つ。

$$f(\mathbf{x} - \mathbf{u}(\mathbf{x}, t), t-1) = f(\mathbf{x} - 2\mathbf{u}(\mathbf{x}, t), t-2) \quad (1)$$

ここで、 $\mathbf{u}(\mathbf{x}, t) = (u(\mathbf{x}, t), v(\mathbf{x}, t))^\top$  はオプティカルフローである [8], [10]。本論文では変分法に基づくオプティカルフロー推定のためのデータ項を次式で定義する。

$$E_D^2(\mathbf{u}, \mathbf{x}, t) = (f(\mathbf{x} - 2\mathbf{u}(\mathbf{x}, t), t-2) - f(\mathbf{x} - \mathbf{u}(\mathbf{x}, t), t-1))^2 \quad (2)$$

式 (2) は不良設定問題であるため、正則化項  $E_S(\mathbf{u}, \mathbf{x}, t)$  を付加する。コスト汎関数は次式で定義される。

$$J(\mathbf{u}, t) = \int_{\Omega} (E_D^2(\mathbf{u}, \mathbf{x}, t) + E_S(\mathbf{u}, \mathbf{x}, t)) dx. \quad (3)$$

本論文では正則化項は次式とする。

$$E_S(\mathbf{u}, \mathbf{x}, t) = \lambda (|\nabla u|^2 + |\nabla v|^2), \quad (4)$$

ここで、正実数  $\lambda$  は正則化項の重みであり、 $\nabla = (\partial_x, \partial_y)^\top$  である。この最小化問題の解  $\mathbf{u}$  は  $\lambda$  の値次第で変化する。変分法より、式 (3) の最小化問題の解は次式のオイラー・ラグランジュ偏微分方程式系の解となる。

$$\begin{aligned} \lambda \nabla^\top \nabla u - E_D \partial_x f &= 0 \\ \lambda \nabla^\top \nabla v - E_D \partial_y f &= 0, \end{aligned} \quad (5)$$

ここで、 $\nabla^\top \nabla = (\partial_x^2 + \partial_y^2)$  である。本論文では斉次ノイマン境界条件とゼロベクトル初期条件を用いる。

マルチ解像度を用いたオプティカルフロー推定のアルゴリズム [8], [11] を図 1 に示す。このアルゴリズムでは

```

入力:  $\lambda$ , 符号化済み  $f(\mathbf{x}, t-2)$  and  $f(\mathbf{x}, t-1)$ ,
最大ピラミッド段数  $l_{\max}$ ,
画素毎の MC のループ回数  $w_{\max}$ ,
逐次緩和 (SOR) 法回数  $k_{\max}$ ,
ピクセル補間法,
動きの初期条件  $\mathbf{u}(\mathbf{x}, t)$ 
出力: 画素毎の動き  $\mathbf{u}(\mathbf{x}, t)$ 
for  $l = 0$  to  $l_{\max}$  do
    build Gaussian pyramids  $f^{[l]}$ 
end for
for  $l := l_{\max}$  to  $0$  do
    for  $w := 1$  to  $w_{\max}$  do
        Compute motion compensated frame  $t$  using  $\mathbf{u}^{[l]}$  and
        decoded frames  $t-2$  and  $t-1$ 
         $\mathbf{u}_{\text{tmp}} := \{\mathbf{0}\} \forall \mathbf{x}$ 
        for  $k := 1$  to  $k_{\max}$  do
            iterate the EL equations for  $\mathbf{u}_{\text{tmp}}$  using frames  $t-2$ 
            and  $t-1$ 
        end for
         $\mathbf{u}^{[l]} := \mathbf{u}^{[l]} + \mathbf{u}_{\text{tmp}}$ 
    end for
    generate  $\mathbf{u}^{[l-1]}$  by up-sampling  $\mathbf{u}^{[l]}$ 
end for
 $\mathbf{u} := \mathbf{u}^{[0]}$ 

```

図 1: 外挿近似のためのオプティカルフロー推定法

$\mathbf{u}_{\text{tmp}}$  は十分に小さいと仮定できるので、 $\mathbf{u}_{\text{tmp}}$  について次式のようにテイラー展開できる [3], [12]。

$$E_D \simeq (\mathbf{u}_{\text{tmp}}^\top \nabla f + \partial_t f). \quad (6)$$

したがって、EL 式は次のようになる。

$$\lambda \nabla^\top \nabla \mathbf{u} - (\nabla f)^\top \nabla f \mathbf{u} + \partial_t f \nabla f = \mathbf{0}, \quad (7)$$

ここで、 $(\nabla f)^\top \nabla f$  は画像の局所構造テンソルであり、 $\mathbf{u}_{\text{tmp}}$  を  $\mathbf{u}$  と略記する。

式 (7) の数値安定性の安定条件は空間刻み幅に依存し、 $\lambda$  と映像に依存しないことが証明されている [13]。したがって、適切な空間刻み幅のもと、任意の  $\lambda$  と映像に対して  $\mathbf{u}$  の数値計算が可能であり、 $\lambda$  の最適化も可能である。式 (5) の数値解を反復計算で求める際、式 (2) の  $E_D^2$  が減少する間、指定した最大反復回数越えなければ反復計算し続ける。ここで、対応する実数点の画素値は線型補間で生成する。

## 3. 参照フレームとしての外挿フレーム

数値計算の結果得られた  $\mathbf{u}$  を用いて  $t-1$  フレームを MC し、 $t$  フレームの近似画像を生成する。この際、対応する  $t-1$  フレームの実数点から Lanczos-4 補間により画素値を生成する。提案法では、近似画像の PSNR が最大となるように、設定した離散的な範囲から最適な  $\lambda$  を求め、復号器側に送信する。復号器では同一の  $\lambda$  と復号済みフレームから同一の  $\mathbf{u}$  を推定できる。

表 1: 参照フレーム 1 枚固定の際の符号量

	外挿フレーム	$t-1$	削減率
Carphone	3.427	<b>3.358</b>	-2.1%
Container	<b>2.800</b>	2.891	3.2%
Foreman	3.587	<b>3.564</b>	-0.6%
Mobile	<b>4.120</b>	4.739	13.1%
News	1.736	<b>1.688</b>	-2.8%
Tempete	<b>4.024</b>	4.327	7.0%
Truck train	<b>3.010</b>	3.259	7.6%
Red leaves	<b>3.645</b>	4.515	19.3%
Fountain	<b>4.098</b>	4.212	2.7%
Studio concert	<b>3.795</b>	3.859	1.7%
Twilight scene zoom	<b>2.910</b>	3.069	5.2%
Horse racing	3.356	<b>3.338</b>	-0.6%
Colorful world	<b>3.679</b>	4.603	20.1%
Average	<b>3.399</b>	3.648	6.8%

表 3: 複数参照フレームの双予測の際の符号量

	$t-1$ から $t-5$ と外挿フレーム	$t-1$ から $t-6$	削減率
Carphone	3.056	3.056	0.0%
Container	2.690	<b>2.686</b>	-0.1%
Foreman	<b>3.163</b>	3.169	0.2%
Mobile	<b>3.911</b>	4.028	2.9%
News	<b>1.590</b>	1.622	2.0%
Tempete	<b>3.776</b>	3.858	2.1%
Truck train	<b>2.589</b>	2.664	2.8%
Red leaves	<b>3.500</b>	3.715	5.8%
Fountain	<b>3.859</b>	3.883	0.6%
Studio concert	<b>3.434</b>	3.474	1.2%
Twilight scene	<b>2.687</b>	2.695	0.3%
Horse racing	<b>3.041</b>	3.178	4.3%
Colorful world	<b>3.393</b>	3.522	3.7%
Average	<b>3.130</b>	3.196	2.1%

表 2: 参照フレーム 2 枚固定の双予測の際の符号量

	$t-1$ と 外挿フレーム	$t-2$ と 外挿フレーム	$t-1$ と $t-2$
Carphone	3.186	3.184	<b>3.144</b>
Container	2.787	<b>2.760</b>	2.770
Foreman	3.359	3.362	<b>3.339</b>
Mobile	4.071	<b>4.030</b>	4.363
News	1.677	1.645	<b>1.642</b>
Tempete	3.943	<b>3.933</b>	4.074
Truck train	2.875	<b>2.848</b>	2.882
Red leaves	3.623	<b>3.573</b>	4.060
Fountain	4.059	<b>4.006</b>	4.045
Studio concert	<b>3.628</b>	3.664	3.659
Twilight scene	2.815	<b>2.765</b>	2.787
Horse racing	<b>3.115</b>	3.319	3.300
Colorful world	3.658	<b>3.601</b>	3.870
Average	3.292	<b>3.284</b>	3.380

提案法では外挿フレームを BM 法の参照ピクチャリストに追加し、従来通りブロック毎の MV を算出する。外挿フレームは等速直線運動を仮定して生成しているため、外挿フレームに対するブロック毎の MV は加速度分のズレを表していることになる。本手法での GOP 構造は IPBBB... とし、外挿フレームは双予測の使える 3 枚目以降を MC 予測する場合に利用できる。したがって、オクルージョン領域や時間的に振動する領域などにおいて外挿フレームの近似精度が低い場合でも、双予測および複数参照フレームを利用することで MC 予測精度の低下を防ぐことができると考えられる。

#### 4. 実験

本実験では、全ての予測ブロックについて MC 予測を利用することで、外挿フレームの本質的な性能を評価する。可変 MC ブロックサイズは  $8 \times 8$  画素から  $32 \times 32$  画

素サイズまでであり、ブロックサイズと整数精度 MV は符号量のもとで最適化する。MV の予測には近傍ブロックの MV からのメディアン予測を用い、予測誤差を算術符号化する。MC 予測誤差はラスタスキャン順に符号化し、近傍画素の予測誤差のコンテキストに基づき適応的に多値算術符号化を行う [14]。I フレームは前置予測のみを用いてイントラ予測する。画像列は 8 ビットグレイスケールの CIF ( $352 \times 288$  画素) サイズである。各画像列 25 フレームの平均符号量を測定する。画素毎の動き推定のための  $\lambda$  は 0.01 から 0.64 までの範囲から 0.01 刻みで最適値を探索し、送信する。

表 1 に、参照フレームを固定して MV を一本だけ送信する MC 予測の場合の符号量を示す。画像列 Mobile、Red leaves、Colorful world では MC 予測精度が向上し、符号量の改善がみられる。これらの画像列では、カメラの移動に伴う局所的な回転や拡大縮小が含まれているため、外挿フレームを参照した方が予測精度が向上すると考えられる。一方、Carphone、Foreman、News では符号量改善は見られない。これは、Carphone と Foreman ではカメラの振動により外挿近似の精度が低下することが原因であると考えられる。さらに、News では、後方のダンサーのオクルージョンにより同様に外挿近似の精度が低下することに加え、画面がほとんど静止しており、外挿の利点が得にくいことが原因であると考えられる。

表 2 に、参照フレーム 2 枚を固定した双予測の場合の符号量を示す。これらの組み合わせのうち、 $t-2$  と外挿フレームを参照する場合が最も効果的であることが分かる。これは、外挿フレームは  $t-1$  フレームの画素値から生成されることと、外挿フレームが  $t-2$  より  $t-1$  に近いためであると考えられる。表 1 同様に、Carphone、Foreman、News では外挿フレームは不利に働いている。

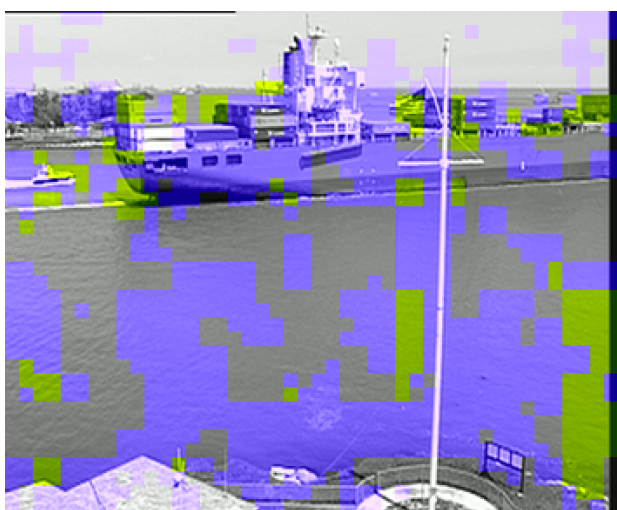
表 3 に、複数参照フレームでの双予測の場合の符号量を



(a) Frame 11 of Carphone



(a) Frame 10 of Red leaves



(b) Frame 13 of Container



(b) Frame 18 of Horse racing



(c) Frame 13 of Foreman



(c) Frame 10 of Colorful world

図 2: 符号量削減率の良くない画像列における外挿フレームが選択されたブロック

示す。この実験では、[15]と同様の手法により6枚の参照ピクチャリストを利用し、7フレーム目の符号化から外挿

図 3: 符号量削減率の良い画像列における外挿フレームが選択されたブロック

フレームを利用する。ブロック毎の参照フレーム選択は符号量の観点から最適化してある。表より、外挿フレームを

用いた複数参照フレーム MC 予測では、外挿フレームを利用しない方法に比べて MC 予測精度を向上させ、総符号量を平均 2%ほど削減できていることがわかる。図 2 と図 3 にそれぞれ符号量削減率の低い画像列と高い画像列での、外挿フレームのブロック適応選択状況を示す。紫の領域は双予測のどちらかで外挿フレームが選択されたブロック、黄緑の領域は双予測の両方で外挿フレームが選択されたブロックを示す。図 2 ではあまり外挿フレームが選択されておらず、外挿近似精度がそれほど高くないことを示している。一方、図 3 ではほとんどのブロックで外挿フレームが選択されているため、外挿近似精度が高いことを示している。さらに、Horse racing での高速に移動する地面領域では特に外挿フレームが選択されている。これは、大きな動きを外挿フレームで動き補償しているため、MV が小さくなり、符号量が削減できていると考えられる。

## 5. おわりに

本研究ではオプティカルフロー推定に基づく外挿近似フレームを用いた複数参照フレーム動き補償予測の性能評価を行った。外挿近似フレームは、微小時間変化での等速直線運動の仮定から、復号器側でわずかな符号量で生成でき、既存のブロック毎の複数参照フレーム MC 予測の枠組みに簡単に追加できる。本手法を用いることで、MC 予測精度が向上し、可逆符号化において符号量を削減できることを示した。ただし、オクルージョンやフレーム間振動のある領域には適していないことも分かった。今後は、非整数精度 MV との比較や、イントラ予測との切り替えなどの組み合わせ、画素毎の動き推定による外挿近似精度の向上について検討する。

## 謝辞

本研究は放送文化基金 (2014 年度~2015) と JSPS 科研費 JP16K00250 の助成を受けたものです。

## 参考文献

- [1] Sullivan, G. J., Ohm, J. R., Han, W. J. and Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1649–1668 (online), DOI: 10.1109/TCSVT.2012.2221191 (2012).
- [2] Sze, V., Budagavi, M. and Sullivan, G. J.(eds.): *High Efficiency Video Coding (HEVC)*, Integrated Circuits and Systems, Springer International Publishing, Cham (2014).
- [3] Brox, T., Bruhn, A., Papenberger, N. and Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping, *European conference on computer vision (ECCV)*, Lecture Notes in Computer Science, Vol. 3024, pp. 25–36 (2004).
- [4] Pock, T., Urschler, M., Zach, C., Beichel, R. and Bischof, H.: A Duality Based Algorithm for TV-L 1-Optical-Flow Image Registration, *Medical Image Computing*

- and *Computer-Assisted Intervention MICCAI 2007*, Vol. 10, No. Pt 2, Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 511–518 (2007).
- [5] Ballester, C., Garrido, L., Lázcano, V. and Caselles, V.: A TV-L1 Optical Flow Method with Occlusion Detection, *Joint 34th DAGM and 36th OAGM Symposium*, pp. 31–40 (2012).
- [6] Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J. and Szeliski, R.: A Database and Evaluation Methodology for Optical Flow, *International Journal of Computer Vision*, Vol. 92, No. 1, pp. 1–31 (online), DOI: 10.1007/s11263-010-0390-2 (2011).
- [7] Jeong, S. G., Lee, C. and Kim, C. S.: Motion-Compensated frame interpolation based on multihypothesis motion estimation and texture optimization, *IEEE Transactions on Image Processing*, Vol. 22, No. 11, pp. 4497–4509 (online), DOI: 10.1109/TIP.2013.2274731 (2013).
- [8] Kameda, Y., Takeichi, J., Ishibashi, M., Matsuda, I. and Itoh, S.: Two stage inter-frame prediction using pixel- and block-wise motion compensation, *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*, No. 2, IEEE, pp. 145–148 (online), DOI: 10.1109/IWSSIP.2015.7314198 (2015).
- [9] Kameda, Y., Takeichi, J., Ishibashi, M., Matsuda, I. and Itoh, S.: Two Stage Motion Compensation Based on Extrapolated Frame by Optical Flow Estimation, *IEICE TRANSACTIONS on Information and Systems (Japanese Edition)*, Vol. J99-D, No. 9, pp. 861–864 (online), DOI: 10.14923/transinfj.2016IEL0006 (2016).
- [10] Horn, B. K. and Schunck, B. G.: Determining Optical Flow, *Artificial Intelligence*, Vol. 17, pp. 185–203 (1981).
- [11] Kameda, Y., Matsuda, I. and Itoh, S.: Numerically stable estimation of scene flow independent of brightness and regularizer weights, *22nd European Signal Processing Conference, {EUSIPCO} 2014, Lisbon, Portugal, September 1-5, 2014*, pp. 1068–1072 (2014).
- [12] Wedel, A., Brox, T., Vaudrey, T., Rabe, C., Franke, U. and Cremers, D.: Stereoscopic Scene Flow Computation for 3D Motion Understanding, *International Journal of Computer Vision*, Vol. 95, No. 1, pp. 29–51 (online), DOI: 10.1007/s11263-010-0404-0 (2011).
- [13] Kameda, Y. and Imiya, A.: Classification of optical flow by constraints, *CAIP (Kropatsch, W. G., Kampel, M. and Hanbury, A., eds.)*, Lecture Notes in Computer Science, Vol. 4673, Springer, pp. 61–68 (2007).
- [14] Matsuda, I., Shirai, N. and Itoh, S.: Lossless Coding Using Predictors and Arithmetic Code Optimized for Each Image, *8th International Workshop, VLBV 2003*, Vol. 5150, pp. 199–207 (2003).
- [15] Maeda, H., Minezawa, A., Matsuda, I. and Itoh, S.: Lossless video coding using multi-frame MC and 3D bi-prediction optimized for each frame, *14th European Signal Processing Conference*, pp. 1–5 (2006).