I-022

複数参照フレームからの DNN ベース幾何変換行列推定を用いたフレーム間予測
Interframe Prediction using DNN-based Geometric Transformation Matrix Estimation from Multiple Reference Frames

姜　思徳†　　八島　由幸†
Side Jiang　Yoshiyuki Yashima

1. はじめに

近年，画像認識を始めとして，カラリゼーションや超解像，画像生成など画像処理の様々な場面にディープラーニングを応用する動きが広がっている．画像圧縮符号化に対して，著者らはこれまでに，畳み込みニューラルネットワーク（CNN）を用いて幾何変換行列を推定し，復号済みフレームに幾何変換を施すことによって，任意精度の平行移動・スケーリングや輝度変化を補償可能なフレーム間予測手法を提案してきた[1]．従来の検討では，過去 2 つの参照フレームから現フレームを生成するフレーム外挿 CNN[1]や時間的に前後の 2 フレームから中間フレームを生成する内挿 CNN[2]の予測性能をそれぞれ評価した．本検討ではこの改良技術として，予測対象となるフレームに対し、内挿 CNN あるいは外挿 CNN を組み合わせ、複数フレームを参照して予測フレームを生成する CNN を構築し，ブロックごとに予測効率の良い CNN を選択する手法を提案する．

2. 提案手法

2.1 変換行列推定に基づくフレーム内挿と外挿

従来より提案している，CNN によって推定された幾何変換行列を用いてフレーム間予測（フレーム内挿とフレーム外挿）を行う手法[1][2]の流れを図 1 に示す．時間的に異なるフレームF, Gから予測対象フレームI中のブロックB_I（サイズ$N \times N$）を予測することを考える．FおよびG中の$\mathbf{B}_I$と同一位置にあるブロックを$\mathbf{B}_F, \mathbf{B}_G$で表す．$\mathbf{B}_I$を中心として，その近傍を拡張した$N_1 \times N_1$サイズのブロックを$\bar{B}_I$とし（$N_1 \geq N$），$\mathbf{B}_F, \mathbf{B}_G$も同様に拡張したブロックを$\bar{\mathbf{B}}_F, \bar{\mathbf{B}}_G$とする．$\bar{\mathbf{B}}_I$の予測値を$\bar{\mathbf{B}}_{ICNN}$とすると，$\bar{\mathbf{B}}_{ICNN}$は$\bar{\mathbf{B}}_F$および$\bar{\mathbf{B}}_G$に CNN によって推定された 4 つの変換行列$\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$を式(1)のように乗算することよって生成される．

$$\bar{\mathbf{B}}_{ICNN} = \mathbf{M}_1(\bar{\mathbf{B}}_F - \mu\mathbf{U})\mathbf{M}_2 + \mathbf{M}_3(\bar{\mathbf{B}}_G - \mu\mathbf{U})\mathbf{M}_4 + \mu\mathbf{U} \quad (1)$$

このとき，$\mathbf{U}$は全ての要素が 1 の行列．μは$\bar{\mathbf{B}}_F$および$\bar{\mathbf{B}}_G$の画素平均値である．学習時の損失計算には$N_2 \times N_2$領域を用い，予測時は$\bar{\mathbf{B}}_{ICNN}$中の$N \times N$領域のブロックを予測値$\boldsymbol{B}_{ICNN}$として使用する．この手法の特徴として，任意精度の平行移動や拡大縮小が可能であり，動きが小さくても局所的に複雑な動きをする動画像に有効であることが確認されている．

† 千葉工業大学大学院情報科学研究科, Graduate School of Information and Computer Science, Chiba Institute of Technology

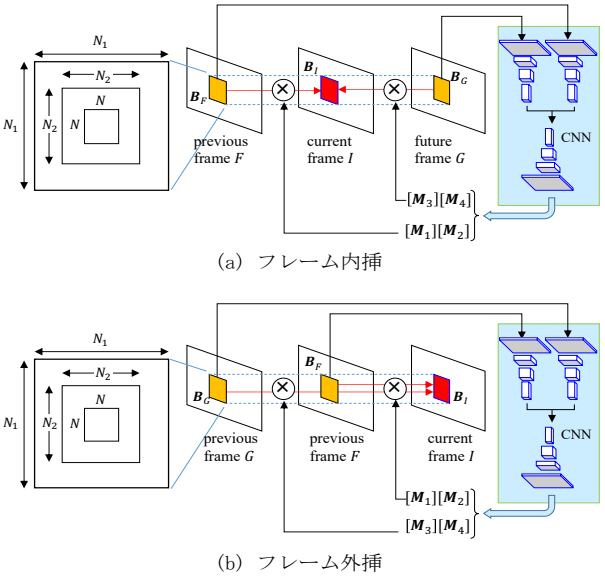

図 1　変換行列によるフレーム間予測

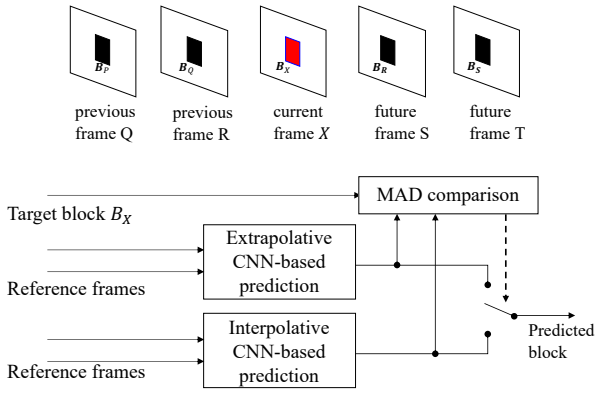

図 2　提案手法

2.2 参照フレームの切り替え

従来の検討ではフレーム内挿予測 CNN やフレーム外挿予測 CNN を個別に検討してきたが，本検討では二つの CNN を組み合わせ，複数フレームを参照して予測フレームを生成する CNN を複数構築し，ブロックごとに予測誤差の少ない CNN を選択する手法を提案する．図 2 に処理の流れを示す．予測対象フレームを X とし，時間的に過去の 2 フレーム Q および R，時間的に未来の 2 フレーム S および T を参照フレームとした場合を対象とする．また，P フレームと B フレームの予測パターンとして，①PBPBPB または②PBBPBB の場合の B フレームを予測対象フレーム

表 1　実験結果（予測画像の MAD）

Sequence	Reference frames : Q, R, S				Reference frames : Q, R, T			
	QR[1]	RS[2]	Average QR&RS	Adaptive QR/RS	QR[1]	RT	Average QR&RT	Adaptive QR/RT
(a)Cosmos flowers	6.701	4.265	4.866	**4.072**	6.701	6.807	6.276	5.788
(b) Sunlight through leaves	7.835	5.725	6.173	**5.429**	7.835	7.219	7.037	6.502
(c) Drama set (day)	1.824	1.643	1.455	**1.253**	1.824	1.92	1.699	1.484
(d) Basketball	4.797	5.398	4.569	**3.492**	4.797	8.298	5.912	4.156
(e) Horse racing (dirt)	4.726	5.604	4.602	**3.674**	4.726	8.298	5.849	4.233

X として想定する．このとき，参照フレームとして用いることができるのは，前者では QRS，後者では QRT となる．

3. 実験と考察

3.1 実験条件

実験では，2.2 節に示した予測パターン①に対して，参照フレームを QR（外挿），RS（内挿），QR と RS の平均(Average)，QR と RS の切り替え(Adaptive)の 4 種類に，予測パターン②に対して，同様に QR, RT, QR と RT の平均，QR と RT の切り替えの 4 種類にした場合の予測フレームと正解フレームとの平均絶対値誤差（MAD）を求めた．ブロックサイズは，$N_1 = 48, N_2 = 32, N = 8$ に固定した．CNN の訓練にはハイビジョン・システム評価用標準動画像第 2 版（960×512 画素に縮小）から 20 種類の動画を用い，評価時には図 3 に示す学習データ以外の 5 種を用いた．Cosmos flowers，Sunlight through leaves，Drama set は動きが少ないが複雑な動きを持つ画像，Basketball と Horse racing は動きの速いカメラワークを伴う画像である．訓練回数は 500 万回，学習率は 0.0001 とした．

3.2 実験結果と考察

表 1 に MAD の測定結果を示す．表 1 より，CNN による予測ブロックごとに予測効率の良い CNN を選択する Adaptive 手法では，動きの大きい動画像に対して有効性を確認することができる．特に QR と RS の組み合わせでは，従来の提案手法と比較すると，動きの小さい動画像でも有効性が見られた．一方，図 4 は QR/RS の適応予測において予測モードの選択割合を動画ごとに示したもの，図 5 は各ブロックがどちらのモードを選択したかを示したものである．これより，動きの大きい場合には外挿予測を選択する割合が多いことがわかる．今回の提案の Adaptive 手法においては，符号化する際に，あらかじめ訓練した CNN をエンコーダとデコーダで共有することで予測を行うことが可能であるが，その際，内挿と外挿の切り替え情報をデコーダに伝送する必要がある．図 5 を見ると，予測モードには空間的な偏りが存在しているため，隣接するブロックと予測モードが等しいことを利用した符号割り当ての工夫によりオーバヘッド符号量の削減が見込める．

4. おわりに

本検討では，複数参照フレームからの CNN を用いたフレーム間予測手法を提案し，その有効性を確認することができた．今後，符号化シミュレーションを行い，復号画質と発生符号量を考慮した RD 最適化に基づく検討を行う予定である．

なお本研究は，JSPS 科研費 JP18K11360 の助成を受けて行った．


(a)Cosmos flowers　(b)Sunlight through leaves　(c)Drama set (day)


(d)Basketball　(e)Horse racing (dirt)

図 3　評価用動画像

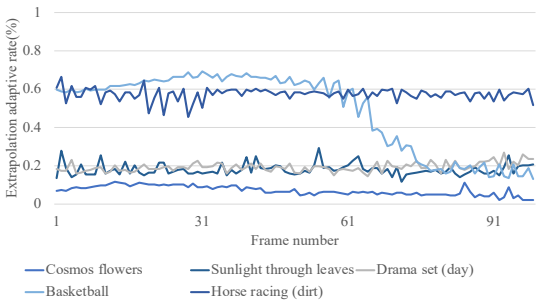

図 4　予測モード選択割合（QR/RS 適応）

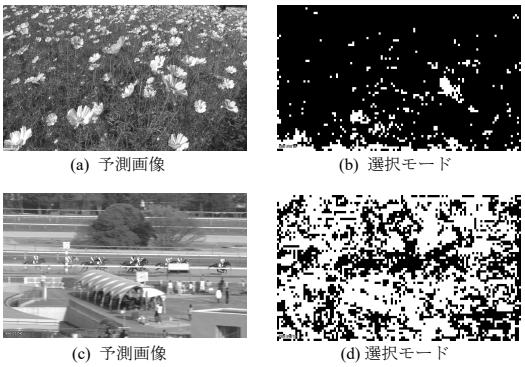
(a) 予測画像　(b) 選択モード
(c) 予測画像　(d) 選択モード

図 5　予測モード（黒：内挿，白：外挿）

参考文献

[1] 神保悟, 王冀, 八島由幸, “深層学習による変換行列予測を用いたフレーム補間,”2017 年映像情報メディア学会年次大会, 2017.
[2] 神保悟，王冀，八島由幸，“深層学習を用いたフレーム間外挿予測と H.265/HEVC への適用,” 信学論， Vol.J102-D, No.10, pp.651-654, 2019.