

# 時空間解像度変換による超高精細映像のリアルタイム符号化

## Real-time Coding of UHDTV with Spatio-temporal Format Conversion

三須 俊枝† 松尾 康孝† 岩村 俊輔† 境田 慎一†  
Toshie Misu Yasutaka Matsuo Shunsuke Iwamura Shinichi Sakaida

### 1. まえがき

8K スーパーハイビジョン (8K) を筆頭として、4K、ハイビジョンなど映像の高解像度化が進行するなか、撮りためてきた映像素材やこれまでに普及してきたテレビ受信機など、素材、伝送、表示の各段階の映像フォーマットにおいてさまざまな解像度・フレームレートの混在が生じることが予想される。

しかし、映像伝送に関しては、ひとたびサービスが開始されると、新たな映像フォーマット登場の度に映像符号化機器等のインフラを変更することには困難が伴う。このため伝送用の映像解像度はそのまま、より高解像度の映像をも伝送できる技術が有用である。

NHK では、超高精細映像の高効率圧縮伝送を実現するため、新しい映像符号化方式である HEVC/H.265 の規格標準化に寄与するとともに、こうした標準的な符号化方式の前後に解像度変換処理を付加することでさらに符号化効率を向上する「超解像復元型映像符号化」方式[1]の研究・開発を推進している。

図1に超解像復元型映像符号化の構成を示す。送信側では入力映像の解像度を削減してから従来型の映像符号化 (コア映像符号化) により圧縮符号化を行う。受信側では、復号された映像を高解像化する。このとき超解像技術を応用した解像度復元を行うことでぼやけの少ない復元映像を得る。また、送信側においては、解像度削減前の原画像を参照しつつ、受信側同様の解像度復元を試行 (ローカル復元) して、その結果が最良のものとなるよう復元法の最適化を行う。最適な復元法を受信側へ補助情報として伝送することで、復元映像の高画質化を可能としている。

これまで、我々は4K 60P映像を2K 60Pに変換し、コア映像符号化 (AVC/H.264) をするシステムにおいてハードウェア実装による実時間処理を実現し、従来型の符号化に比べて高画質化が可能であることを確認してきた[2]。しかしこの時点では、解像度変換は空間解像度のみに限られ、また、送信側の最適化処理については未実装であった。

今回、空間解像度に加えて時間解像度 (フレームレート) の変換と最適化処理を実装した実時間解像度変換装置 (図2) を開発し、コア映像符号化としての HEVC/H.265 コーデックと組み合わせたシステムを構築した。加えて、今回のシステムでは、表1に示す解像度削減と解像度復元を任意に組み合わせることも可能とすることで、映像素材とテレビ画面のさまざまな解像度の混在にも対応できるマルチフォーマット映像符号化を実現した。

### 2. システム構成

超解像復元型映像符号化では、送信側においてローカル復元を行うことから、送信側には受信側の機能を包含する必要がある。そこで、今回試作した時空間解像度変換装置は送受信を別々に設計するのではなく、送受信で兼用できる構成をとることとした。

時空間解像度変換による超解像復元型映像符号化の構成を図3に示す。時空間解像度変換装置は、時空間解像度削減部と、時空間ハイブリッド復元部と、最適化・制御部で構成される。

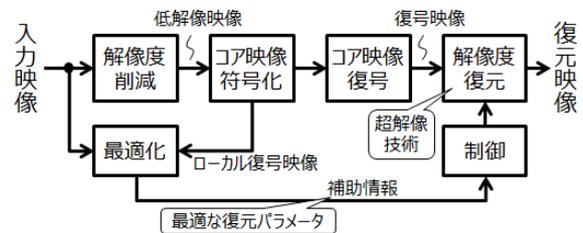


図1. 超解像復元型映像符号化の構成



図2. 時空間解像度変換装置 (19吋 EIA 4U サイズ)

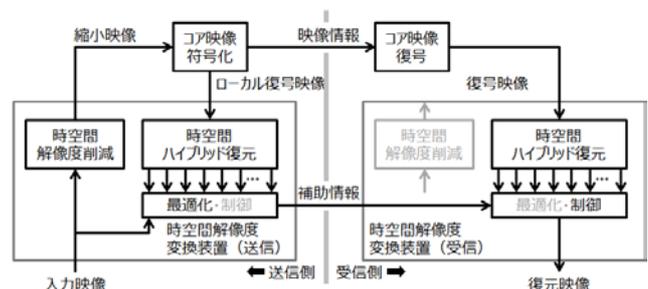


図3. 時空間解像度変換による超解像復元型映像符号化

表1. 解像度変換装置のサポートする解像度変換

出力 入力	2K 60P	2K 120P	4K 60P	4K 120P
2K 60P		時間復元	空間復元	時空間復元
2K 120P	時間削減		-	-
4K 60P	空間削減	-		時間復元
4K 120P	時空間削減	-	時間削減	

†日本放送協会, NHK

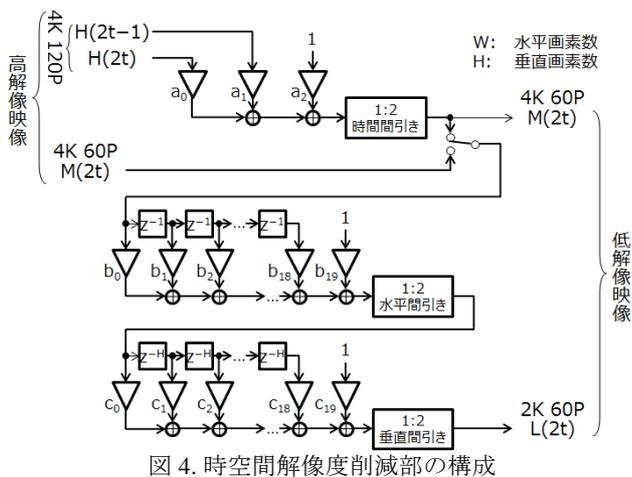


図 4. 時空間解像度削減部の構成

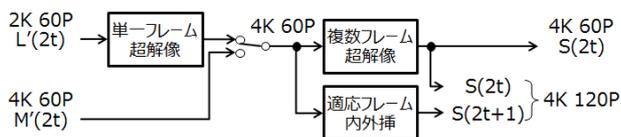


図 5. 時空間ハイブリッド復元部

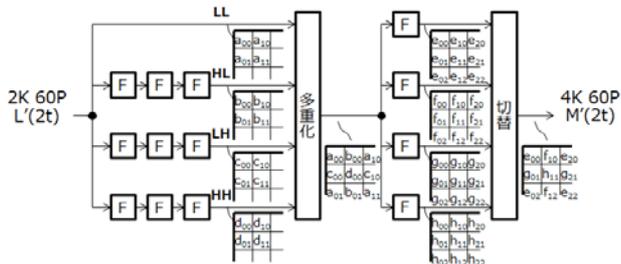


図 6. 単一フレーム超解像部

2.1 時空間解像度削減部

時空間解像度削減部では、時間解像度削減の後、空間解像度削減を行う。図 4 に示すように、両削減ともに、間引きフィルタを適用してから標本点を間引くことで、折り返し歪を適切にカットする（または、付加する）ことができるよう構成されている。以降の説明では、入力される 4K 120P 高解像映像を  $H(T,x,y)$  ( $T$  は 120Hz におけるフレーム番号,  $(x,y)$  は画像座標) とおく。なお、高解像映像入力が 4K 60P 映像の場合には  $M(2t,x,y)$  ( $t$  は 60Hz におけるフレーム番号) と表記する。

時間間引きフィルタは 2 つのフレームの各画素位置に対して、2 タップの畳み込みとオフセット付加を行う計 3 タップのデジタルフィルタである。時間間引きは、2 フレームにつき 1 フレームだけをピックアップして出力する。

空間解像度削減は、まず水平方向に 19 タップの係数を畳み込み、水平方向に 1/2 の間引きを行う。続いて、垂直方向

に 19 タップの係数を畳み込み、垂直方向に 1/2 の間引きを行う。

時間解像度削減や空間解像度削減をバイパスすることができる構成し、時間のみまたは空間のみの削減も可能とした。

表 2. 動き推定のパターン

インデックス $k$	推定元 $a_k$	推定先 $b_k$	補償元 $c_k$	補償先 $d_k$	按分 $p_k$
0 (超解像)	-4	-8	-8	-4	/
1 (超解像)	-4	-6	-6	-4	
2 (超解像)	-4	-2	-2	-4	
3 (超解像)	-4	0	0	-4	
4 (外挿)	-2	0	-2	-3	-0.5
5 (内挿)	-4	-2	-4	-3	+0.5
6 (内挿)	-2	-4	-2	-3	+0.5
7 (外挿)	-4	-6	-4	-3	-0.5

2.2 時空間ハイブリッド復元部

時空間ハイブリッド復元部は、図 5 に示すように、

- 単一フレーム超解像
- 複数フレーム超解像
- 適応フレーム内外挿

の 3 つの復元手法により構成した。

(1) 単一フレーム超解像部

単一フレーム超解像部にはウェーブレット超解像方式を採用した。本方式では、図 6 のように入力映像の高域成分から出力すべき高解像映像の高域成分 (HL, LH, および HH の各成分) を生成し、入力映像を LL 成分とみなしてウェーブレット合成を行うことで 2K 60P から 4K 60P への空間超解像を行う。図 6 の共通演算ユニット F は、いずれも  $5 \times 5$  タップの 2 次元畳み込みとオフセットからなる 26 タップのデジタルフィルタである。共通ユニットの並・直列配置は、ハードウェア設計の効率化や回路の高集積化に有利な構造である。

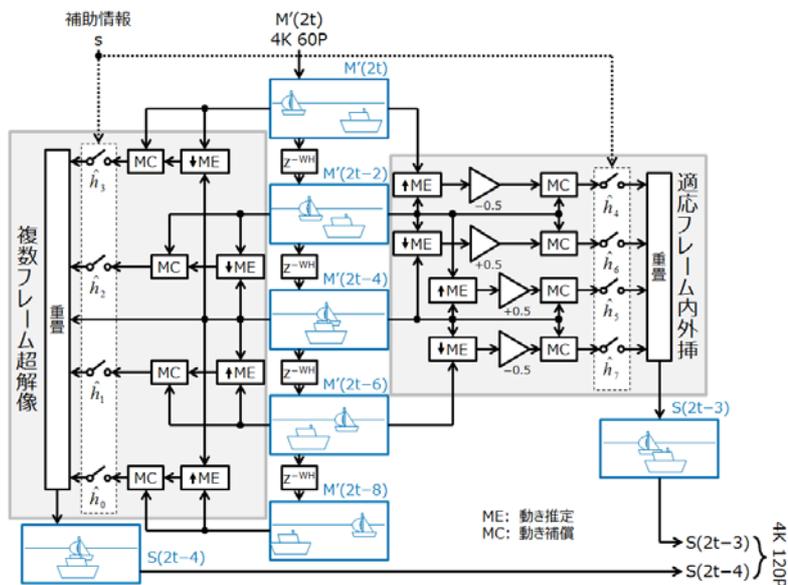


図 7. 複数フレーム超解像部および適応フレーム内外挿部

## (2) 複数フレーム超解像部

複数フレーム超解像部と適応フレーム内外挿部は、図 7 に示すように単一フレーム超解像の結果を過去 5 フレーム分保持するフレームバッファと動きベクトル演算を共通化することにより、ハードウェアリソースを節約している。以下、フレームバッファに記憶されたフレームの時刻列を $(-8, -6, -4, -2, 0)$  (最新フレームの時刻を 0) と定義し、時刻  $T$  における単一フレーム超解像の結果の画像座標 $(x,y)$ における画素値 $M'(T, x, y)$ とおく。

動きベクトル演算においては、時刻  $a$  のフレーム  $M(a, x, y)$  上に設定したブロック  $B$  に対応する時刻  $b$  のフレーム  $M(b, x, y)$  上の領域を探索し、動きベクトル $(u_{a,b}, v_{a,b})$  を求める。

$$(u_{a,b}, v_{a,b}) = \arg \min_{(u,v)} \sum_{(x,y) \in B} |M'(2t+b, x+u, y+v) - M'(2t+a, x, y)|$$

複数フレーム超解像部および適応フレーム内外挿部において動き推定および動き補償のために参照するフレーム  $a, b, c$ 、および  $d$  の組み合わせを表 2 に示す。

複数フレーム超解像部は、まず、  
 $J_k(x, y) = M'(2t+c_k, x+u_{-4,b_k}, y+v_{-4,b_k})$  for  $(x, y) \in B$   
 により、4 通り ( $k \in \{0,1,2,3\}$ ) の動き補償画像  $J_k$  を生成する。続いて、これらの 4 つの画像の使用/不使用を表す重畳フラグ  $\hat{h}_k$  に従って画像重畳を行い、超解像結果  $S$  を得る。

$$S(2t-4, x, y)|_{(x,y) \in B} = \sum_{k=0}^3 \hat{h}_k(B) J_k(x, y) / \sum_{k=0}^3 \hat{h}_k(B)$$

## (3) 適応フレーム内外挿部[3]

適応フレーム内外挿部は、動きベクトルを按分した動き補償

$$F_k(B; x+p_k u_{a_k, b_k}, y+p_k v_{a_k, b_k}) = \begin{cases} M'(2t+c_k, x, y), & (x, y) \in B \\ 0, & (x, y) \notin B \end{cases}$$

$$G_k(B; x+p_k u_{a_k, b_k}, y+p_k v_{a_k, b_k}) = \begin{cases} 1, & (x, y) \in B \\ 0, & (x, y) \notin B \end{cases}$$

により、各ブロック  $B$  に対して 4 通り ( $k \in \{4,5,6,7\}$ ) の動き補償画像  $F_k$  およびマスク画像  $G_k$  を生成する。これらを、重畳フラグ  $\hat{h}_k$  に応じて次式によって重畳することに

より内外挿結果  $S$  を得る。なお、 $\sum_{k=4}^7 \hat{h}_k(B) G_k(B; x, y) = 0$  なる領域 (動き補償の結果が重ならず穴が空いてしまった領域) に対しては時刻 $-2$  および $-4$  の単一フレーム超解像結果の画素ごとの平均値 (すなわち線形内挿) の結果をはめ込むものとする。

$$S(2t-3, x, y) = \sum_{B \in \text{全ブロック}} U(B; x, y)$$

$$U(B; x, y) = \begin{cases} \frac{M'(2t-2, x, y) + M'(2t-4, x, y)}{2}, & \sum_{k=4}^7 \hat{h}_k(B) G_k(B; x, y) = 0 \\ \sum_{k=4}^7 \hat{h}_k(B) F_k(B; x, y) / \sum_{k=4}^7 \hat{h}_k(B) G_k(B; x, y), & \text{otherwise} \end{cases}$$

## 2.3 最適化・制御部

最適化機能は、送信側にも時空間ハイブリッド復元を設け、ローカル復号映像を表 2 の各パターンの採否をさまざまに変えて解像度復元を試行し、ブロック  $B$  ごとに原画像と比較を行うことで補助情報を生成する機能である。

複数フレーム超解像の最適化は、重畳部において動き補償画像  $J_k$  を重畳するか否かの全パターン 16 通りを試行し、得られた超解像結果が高解像映像  $H(2t-4)$  を最も近似する重畳フラグ  $\hat{h}_k$  を、

$$\begin{bmatrix} \hat{h}_0(B) \\ \hat{h}_1(B) \\ \hat{h}_2(B) \\ \hat{h}_3(B) \end{bmatrix} = \arg \min_{\substack{\sum_{(x,y) \in B} \sum_{k=0}^3 h_k J_k(x, y) \\ \sum_{k=0}^3 h_k \\ h_i \in \{0,1\}^4}} \left| \frac{\sum_{k=0}^3 h_k J_k(x, y)}{\sum_{k=0}^3 h_k} - H(2t-4, x, y) \right|$$

により求めて、補助情報として出力する。

適応フレーム内外挿の最適化は、動き補償画像  $F_k$  ( $k \in \{4,5,6,7\}$ ) と高解像映像  $H(2t-3)$  を比較し、その差分絶対値和 (SAD) が閾値  $\theta$  以下であれば重畳ありと判定する。

$$\hat{h}_k(B) = \begin{cases} 1, & e_k(B) \leq \theta \\ 0, & \text{otherwise} \end{cases}$$

$$e_k(B) = \sum_{(x,y) \in B} \left| \frac{F_k(B; x+p_k u_{a_k, b_k}, y+p_k v_{a_k, b_k})}{-H(2t-3, x+p_k u_{a_k, b_k}, y+p_k v_{a_k, b_k})} \right|$$

for  $k \in \{4,5,6,7\}$

こうして得られたブロック  $B$  ごとの重畳フラグ  $\hat{h}_k(B)$  ( $k \in \{0,1,\dots,7\}$ ) は補助情報として送信側から受信側へと伝送される。以下では、重畳フラグ  $\hat{h}_k(B)$  を 2 進数の LSB 側から第  $k$  ビットに配することにより、ブロック  $B$  ごとの補助情報  $s(B)$  を 8 ビット値として表現する。

$$s(B) = \sum_{k=0}^7 2^k \hat{h}_k(B)$$

なお、送信側において空間解像度のみの削減 (4K 60P 素材を 2K 60P に変換するなど) を行った場合には重畳フラグ  $\hat{h}_k(B)$  のうち  $k \in \{0,1,2,3\}$  なるもののみを伝送する。逆に、時間解像度のみの削減 (2K 120P 映像素材を 2K 60P に変換するなど) を行った場合には重畳フラグ  $\hat{h}_k(B)$  のうち  $k \in \{4,5,6,7\}$  なるもののみを伝送する。

受信側の制御部においては、受信した補助情報  $s(B)$  に従って重畳フラグ  $\hat{h}_k(B)$  の 0/1 を設定し、複数フレーム超解像部および適応フレーム内外挿部を制御する。

このとき、重畳フラグ  $\hat{h}_k(B)$  の一部が利用できない場合 (送信側で空間のみまたは時間のみ削減を行った場合) には、時空間ハイブリッド復元部は、重畳フラグの得られない処理についてはデフォルトのパラメータにより時空間復元を行う。

## 2.4 信号同期部

図 2 に示したシステムにおいては、以下の 2 箇所において信号同期が必要となる。

- 送信同期： 高解像映像とローカル復号映像の同期
- 受信同期： 復号映像と補助情報の同期

これらの信号同期を実現するため、今回開発した時空間解像度変換装置においては、

- マーカ方式 (送受両用, 映像信号に対し侵襲的)
  - ハッシュ方式 (受信専用, 映像信号に対し非侵襲的)
- の 2 種類を実装した。

## (1) マーカ方式

マーカ方式では、低解像映像  $L$  の左上  $32 \times 32$  画素に白または黒のベタ塗りを 5 秒周期で交互に重畳するとともに、いずれのベタ塗りが表示されているかを示すフラグを補助情報として伝送する。送信同期は、低解像映像  $L$  に重畳したマーカの切り替えのタイミングとローカル復号映像上において観測されたマーカの切り替えのタイミングから映像符号化において生じる遅延量を求め、最適化部において参照される高解像映像  $H$  のタイミングをとる。受信同期用には、復号映像  $L$  において観測されたマーカ切り替えのタイミングと、補助情報から得たフラグ変化のタイミングに基づき復号映像と補助情報の同期をとる。

## (2) ハッシュ方式

送信側のローカル復号映像と、受信側の復号映像が信号レベルで一致することを利用して、映像情報ストリームと補助情報ストリームの同期をとる独自の方式である。送信側でローカル復号映像の各フレームを特徴づける数値(ハッシュ値、今回は 16 ビット値)を求め、補助情報に多重化して伝送する。受信側でも、復号映像からハッシュ値を求め、補助情報中のハッシュ値と対応付けを行うことで、同期をとる。

本方式は、映像情報には一切の同期情報を付加していないことが最大の長特である。このため、コア映像符号化がタイムコード情報を通さないような系においても信号同期を実現することができる。



図 8. 時空間解像度変換装置を用いた映像符号化の実演

## 3. 実験結果

4K 120P 映像(4K 60P 映像素材を倍速再生)を時空間解像度変換することで 2K 60P 用 HEVC エンコーダで符号化する実験を行った。HEVC によるコア映像符号化における符号化遅延に加え、わずか 0.18 秒程度の低遅延を実現した。図 8 は、NHK の「技研公開 2014」において 4K 120P 映像を 16Mbps (映像 12Mbps, 補助情報 4Mbps) により符号化した実演のようすである。

図 9, 図 10 は 4K 120P 映像を 2K 60P 用 HEVC エンコーダ 8 並列 (2Mbps  $\times$  8) により符号化した場合と提案法の画質比較である。各図(a)に示す 8 並列での符号化結果は、とくに動きの異なる被写体が重なり合う箇所において顕著な劣化(図 9(a)の○印)が見られたが、提案手法では、図 9(b)の複数フレーム超解像結果、図 10(b)の適応フレーム内外挿の結果ともに画質が改善している。一方で、静止画に近い映像においては HEVC エンコーダ 8 並列の方がやや精細感があった(図 10(a)の□印)。

提案手法は 4 つの動き補償結果を適応的に選択・重畳する仕組みを備え、複数の動きの交錯するシーンに有利な

のに対し、静止画に近い領域では解像度削減での損失が解像度復元で完全には回復できなかったことに起因すると考えられる。

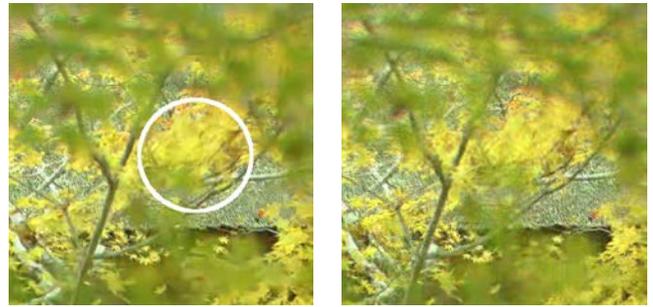


図 9. 画質比較 (偶数フレーム)



図 10. 画質比較 (奇数フレーム)

## 4. まとめ

時空間解像度変換装置を試作し、HEVC エンコーダとの組み合わせによって 4K 120P 映像の実時間映像符号化を実現した。

本符号化方式により、伝送用の映像解像度を変えずに、より高解像のさまざまな映像素材を伝送することや、多様な解像度のディスプレイに提示することも可能となる。

現時点では、本方式による画質の客観的/主観的評価が不足している。今後、評価実験行いつつ復元処理のパラメータや補助情報の割合を最適化することで高画質化を図るとともに、8K スーパーハイビジョンの解像度に対応させる予定である。

## 参考文献

- [1] T. Misu, Y. Matsuo, S. Sakaida, Y. Shishikui, and E. Nakasu, "Novel Video Coding Paradigm with Reduction/Restoration Processes," *Proceedings of PCS 2010*, S2-2, pp.466-469, December 2010.
- [2] T. Misu, Y. Matsuo, S. Iwamura, and S. Sakaida, "Real-time Implementation of UHDTV Video Coding System with Super-resolution Techniques," *Proceedings of PCS 2013*, pp. 181-184, May 2013.
- [3] 三須, 松尾, 岩村, 井口, 境田, "複数フレーム参照によるフレーム内外挿法とその映像符号化への応用," *FIT 2012*, No.3, I-019, pp.289-292, September 2012.