

超解像復元技術を用いる 4K 12 ビット実時間映像符号化システム 4K 12-bit Real-time Video Coding System with Super-resolution Techniques

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 (以下, 8K) の実用化を目前に控え, 放送, 素材伝送などを想定したさまざまな映像伝送システムの開発・検証が急ピッチで進められている. 8K 映像は, 最大で 7680×4320 画素, フレームレート 120Hz, ビット深度 12 ビット, BT.2020 色域, クロマサンプリング 4:4:4 と規定されるが, とくに初期の段階や移行期においては上記スペックを扱うに足りない帯域や処理能力の符号化・伝送システム上においても満足いく品質で映像を伝送できることが必要となる.

これまで筆者らは, 超解像技術による 4K/2K の変換や 120Hz/60Hz の変換, パラメータ最適化の機能を従来の映像符号化方式とともに組み合わせた超解像復元型映像符号化システムを開発し, とくに低ビットレート時において顕著な画質改善が可能であることを示してきた[1].

いずれも従来の符号化技術を改変することなく利用し, その外側に処理を追加することで符号化効率向上や対象映像フォーマットの拡張を図っている. 本システムでは, 従来符号化の前には解像度等の削減処理を設け, 従来符号化における圧縮率が極端に高くなることを回避することで符号化劣化を抑制する. 一方, 従来復号の後には解像度等の復元処理を設ける. ここに超解像技術を用いることで, 解像度復元時のぼやけを防いでいる.

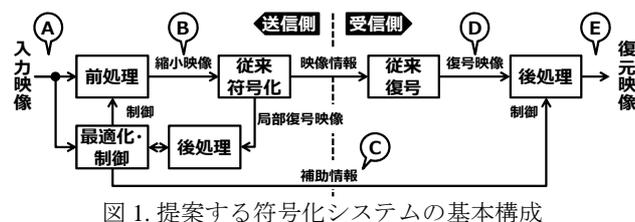


図 1. 提案する符号化システムの基本構成

2. システム構成

図 1 に示すように, 本システムは従来型の映像符号化・復号の前後に処理を追加した形態を基本とする. また, 送信側においては, 受信側の後処理を模擬した結果が入力画像に近づくよう最適化を行う. 最適化結果は補助情報として受信側に伝送し, 後処理を制御する.

これまで, 筆者らは送信側前処理/受信側後処理のペアとして,

- ① 空間解像度削減/空間解像度復元
- ② 時間解像度削減/時間解像度復元
- ③ 時空間解像度削減/時空間解像度復元
- ④ 空間解像度・階調削減/空間解像度・階調復元

などを検証してきた. これらの復元処理には削減処理時における高域のロスを補う超解像技術を導入し, ぼやけの少ない復元を実現している. また, とくに①と③については

FPGA ベースの映像処理装置を開発し, 最大で 4K 解像度, フレームレート 120Hz における実時間処理を達成している [1].

今回は, ④の空間解像度および階調の削減/復元の実時間処理装置を開発し, 最大で 4K 解像度 12 ビットの映像を 2K 解像度 8 ビット用の従来符号化装置で実時間伝送を可能とした. また, 誤差のほか構造的なものを考慮した規範により最適化処理を高度化した. 映像情報と補助情報の自動調相機能を実装し, 両者が別経路で送られる場合にも同期が可能であることを実証した.

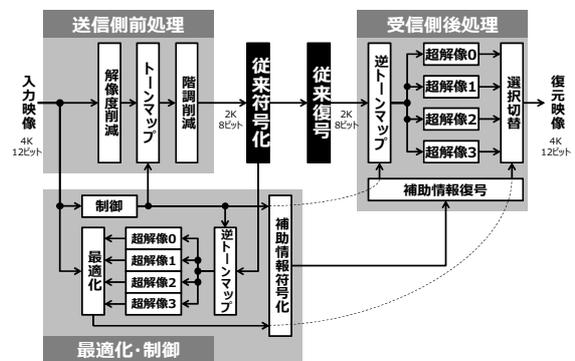


図 2. 空間解像度・階調変換に基づく映像符号化システム

2.1 空間解像度・階調変換

今回開発した空間解像度・階調変換に基づく映像符号化システムの構成を図 2 に示す.

送信側前処理においては, 空間解像度を水平・垂直ともに 1/2 に変換する解像度削減の後, 階調を非線形変換するトーンマップを適用し, 階調数を削減する.

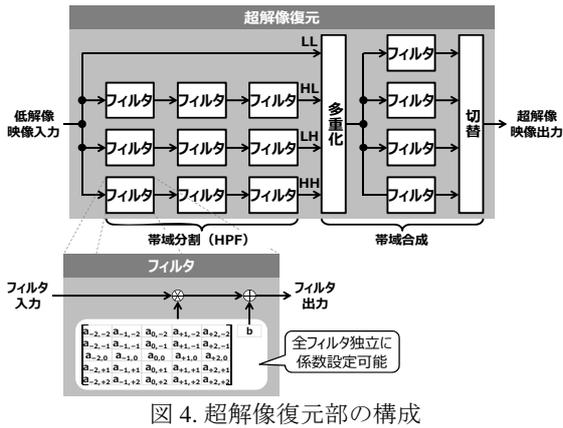
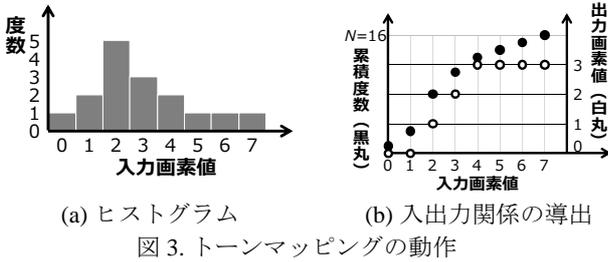
受信側後処理においては, 前処理におけるトーンマップの逆変換である逆トーンマップを適用後, 空間解像度を水平・垂直ともに 2 倍に超解像復元する. このとき, 高域付加の強度など画質の異なる復元処理を 4 通り実行し, それらの結果をフレーム・色・ブロック単位で選択切替できる構成されている.

さらに送信側には, 受信側と同様の後処理 (逆トーンマップおよび超解像復元) が実装されている. 従来符号化による局部復号映像に対してこれらの後処理を適用し, その結果は入力映像と比較される. 4 つの超解像復元の結果のうちいずれが入力映像に近いかをフレーム・色・ブロック単位で判定し, その判定結果を補助情報として受信側に伝送する. この補助情報伝送により, 受信側においても常に最適な超解像復元結果を選択可能となる.

2.2 空間解像度削減

空間解像度削減部では, 4K 解像度の入力映像を 2K 解像度にダウンコンバートする. このとき入力映像の各色成分に対し, それぞれ 17×17 タップの間引きフィルタを適用

後(係数は色毎に設定可能), 水平・垂直ともに 2 画素単位で画素を間引く処理が行われる。



2.3 トーンマップ・階調削減

トーンマップ部は画素値をルックアップテーブル (LUT) によって変換することで, 入出力間において階調を非線形的に変換する。階調削減処理による情報のロスを極力軽減するため, トーンマップ部は後段の階調削減部とともに, 出現頻度の高い画素値付近ほど細かく再量子化する。

このとき, 入力映像フレームの各色について, 画素値のヒストグラムが平坦化するよう処理を行うこととした。さらに今回の実装では, 画像を上下左右に 4 分割し, 各部分領域に対して独立にトーンマッピングを施すことで, 画像の局所的な階調性の差異に対応した。

図 3 にトーンマップ部の動作を例示する。まず, 入力映像フレームの各色成分(輝度, 色差)の, 上下左右にそれぞれ 2 分割した各領域につき画素値のヒストグラムを作成する(図 3(a))。続いて, このヒストグラムを累積することで累積ヒストグラムを生成する(図 3(b)の黒丸)。累積ヒストグラムの最大値は入力映像フレーム 1/4 領域の総画素数 N となる。ここで, トーンマッピングおよび階調削減による出力階調数を Q 階調としたい場合, 累積ヒストグラムの縦軸を Q 個に等分するよう閾値を設定する。

ここで, 累積度数が閾値 qN/Q を超え, 閾値 $(q+1)N/Q$ 以下にある場合には, 出力画素値を q とすることでトーンマッピングの入出力関係を定義する(図 3(b)の白丸)。このような手順をとることにより, 変換後の画像においては, 階調数が削減されるとともに, そのヒストグラムが平坦化される。

なお, 入出力間の変換には LUT を用い, 毎フレーム, 1/4 画面単位で LUT が更新できるよう構成した。

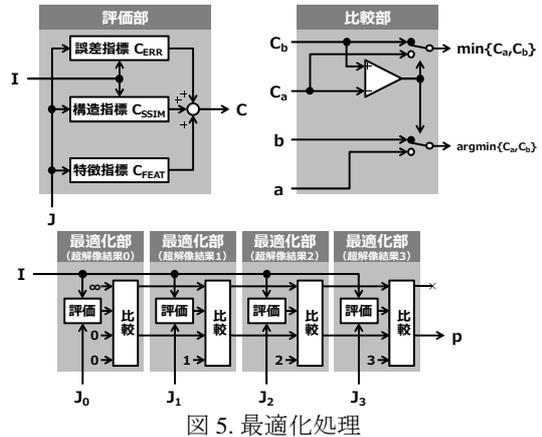
2.4 超解像復元

本システムにはウェーブレット超解像方式による超解像復元部が 4 式実装され, それぞれパラメータ設定値によって超解像の度合いや質の異なるアップサンプリングを行う。これらをフレーム・色・ブロック単位で適宜切り替えることによって適応的なアップサンプリングを実現する。

各超解像復元部の構成を図 4 に示す。超解像復元部は, 帯域分割部と帯域合成部に大別される。帯域分割部では, 低解像入力映像に対し, 水平, 垂直および対角方向の高域通過型フィルタ (HPF) を適用することで, 垂直, 水平および対角エッジを生成する。

低解像入力映像と, これらエッジ画像とからなる 4 バンドの画像を帯域合成部において合成することにより水平, 垂直それぞれ 2 倍の画像を生成する。この帯域合成は Haar ウェーブレット等の基底関数により実行される。

帯域分割には 9 種類 (3 段×3 帯域), 帯域合成には 4 種類のデジタルフィルタが実装されている。いずれも 5×5 の 2 次元畳込演算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。



2.5 最適化処理

受信側超解像復元処理の 4 択を制御するため, 送信側において入力映像を参照しつつ最適化を行う。最適化は, 図 5 に示すように 4 つの復元結果 $J_0(x,y) \sim J_3(x,y)$ を逐次比較する構成とした。

入力映像フレームを $I(x,y)$, 超解像復元 n (図 2 では超解像 n と表記) による復元結果を $J_n(x,y)$, 出力する復元映像フレームを $S(x,y)$ とおく。

このとき, 復元映像 S は, ブロック B 単位で復元映像 $J_0(x,y) \sim J_3(x,y)$ を切り替えることにより生成される。今回のシステムではブロック B を 2K 映像上において輝度は 32×32 画素, 色差は 16×32 画素のサイズとした (クロマサブサンプリングを 4:2:2 としたため)。

$$S(x,y)|_{(x,y) \in B} = J_{p(B)}(x,y)$$

ブロック B ごとにいずれの超解像復元結果を用いるべきかを判定した結果を $p(B) \in \{0,1,2,3\}$ とおく (以下, フレーム内に存在する $p(B)$ の列を切替マップと呼ぶ)。

人間の視覚系による主観画質に近い判定を実現するため, 切替マップ $p(B)$ の導出には, 誤差指標 C_{ERR} , 構造指標 C_{SSIM} および特徴指標 C_{FEAT} を組み合わせた最適化コスト汎関数 $C[J_n](B)$ を用いる。なお, 最適化コスト汎関数

$C[J, I](B)$ は、映像信号波形 I および J をブロック B 内で評価し、その結果をスカラー値で出力するものとする。なお、誤差指標と構造指標は復元映像フレーム J_n と入力映像 (リファレンス) I との両者によって定義したのに対し、特徴指標は復元映像 J のみによって定まるものとした。すなわち、

$$C[J, I](B) = C_{ERR}[J, I](B) + C_{SSIM}[J, I](B) + C_{FEAT}[J](B)$$

とおくことで、切替マップ $p(B)$ を、

$$p(B) = \arg \min_{n \in \{0,1,2,3\}} \{C[J_n, I](B)\}$$

により得る。

誤差指標 C_{ERR} は、

$$C_{ERR}[J, I](B) = L_2 \left(\sum_{(x,y) \in B} L_1(I(x,y) - J(x,y)) \right)$$

と定義した。 L_1 および L_2 は LUT であり、テーブルの中身は自由に設定することができる。 L_1 および L_2 の設定次第で、RMSE (Root Mean Squared Error) や SAE (Sum of Absolute Errors) により誤差評価を行うことができる。例えば、

$$L_1(x) = x^2 / |B|, \quad L_2(x) = \sqrt{x}$$

のように LUT を設定すれば RMSE を評価することができ、

$$L_1(x) = |x| / |B|, \quad L_2(x) = x$$

とすれば SAE を評価することができる。なお、 $|B|$ はブロック内の総画素数である。このように LUT の設定次第でさまざまなノルムでの誤差評価が可能である。

構造指標 C_{SSIM} は構造類似性指標 (SSIM; Structural Similarity) [2] を評価する指標であり、以下のように定義される。

$$C_{SSIM}[J, I](B) = A \frac{(2\mu_I(B)\mu_J(B) + c_1)(2\sigma_{IJ}(B) + c_2)}{(\mu_I^2(B) + \mu_J^2(B) + c_1)(\sigma_I^2(B) + \sigma_J^2(B) + c_2)}$$

$$\mu_I(B) = \sum_{(x,y) \in B} I(x,y) / |B|, \quad \mu_J(B) = \sum_{(x,y) \in B} J(x,y) / |B|$$

$$\sigma_I^2(B) = \sum_{(x,y) \in B} (I(x,y) - \mu_I)^2 / |B|$$

$$\sigma_J^2(B) = \sum_{(x,y) \in B} (J(x,y) - \mu_J)^2 / |B|$$

$$\sigma_{IJ}(B) = \sum_{(x,y) \in B} (I(x,y) - \mu_I)(J(x,y) - \mu_J) / |B|$$

SSIM のパラメータ c_1 および c_2 は自由に設定可能である。

また、スカラー A は他の指標とのバランスを調整するための係数であり、これも自由に設定可能である。

特徴指標 C_{FEAT} は、畳み込み (係数は L_3 および L_5 の LUT に保存) および関数 (L_4 および L_6 の LUT で表現) を用いて以下のように定義する。

$$C_{FEAT}[J_n](B) = \sum_{(x,y) \in B} \left\{ L_4 \left(\sum_{j=-1}^{+1} \sum_{i=-1}^{+1} L_3(i,j) J_n(x+i, y+j) \right) + L_6 \left(\sum_{j=-1}^{+1} \sum_{i=-1}^{+1} L_5(i,j) J_n(x+i, y+j) \right) \right\}$$

このようにすることで、例えば、

$$L_3(i,j) = \begin{cases} -1, & i = -1, j = 0 \\ +1, & i = 0, j = 0 \\ 0, & \text{otherwise} \end{cases}, \quad L_5(i,j) = \begin{cases} -1, & i = 0, j = -1 \\ +1, & i = 0, j = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$L_4(x) = L_6(x) = |x|$$

なる LUT を設定することで、トータルバリエーション

(TV) 値を求めることができる。TV 値は信号波形が振動的かどうかを判定するのに有効であり、ウェーブレット超解像が失敗した際に発生しがちなリングングの検出に寄与する。

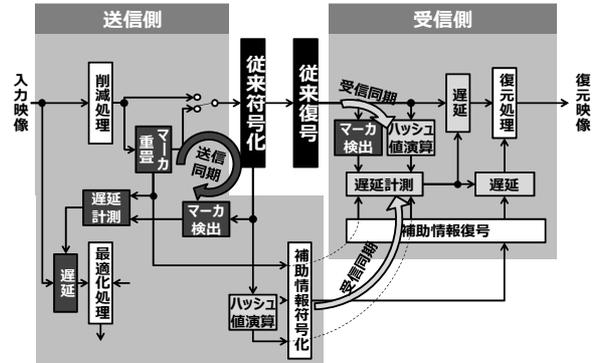


図 6. 信号同期系

2.6 信号同期

本システムでは、従来の映像符号化装置の出力するストリームと、補助情報符号化装置の出力する補助情報ストリームの 2 種類のデータを伝送する必要がある。両ストリームは、多重化して送ることも別々に送ることも想定している。しかし、とくに映像ストリームと補助情報ストリームを別々に送る場合、両ストリームの同期を確立する必要がある。このため、今回開発した装置には図 6 に示すような信号同期機能を搭載している。

信号同期のためにはタイムスタンプを両ストリームに付与するのが一般的である。しかし、従来型の映像符号化装置や伝送装置の中にはタイムスタンプ等のアンシラリデータを伝送できないものもあることから、映像ストリーム側に細工をすることは好ましくない。

そこで、今回開発した空間解像度・階調変換装置では、送信側における最適化処理に併行して、現在処理している局部復号映像フレームの映像特徴量 (ハッシュ値) を計算し、切替マップ $p(B)$ とともに多重化して補助情報伝送する「ハッシュ同期方式」を採用した。

ハッシュ同期方式において、受信側ではこれから復元処理する復号映像フレームのハッシュ値を求める。これと同じハッシュ値を有する補助情報を探索することで、正しいタイミングの切替マップ $p(B)$ を得ることができる。

今回は、ハッシュ値として、画像フレームの上部 32 ラインの輝度値総和の LSB 側 16 ビットを用いている。この場合、ごく稀に異なる映像フレームに対して同一のハッシュ値が出現することもあるが、

- 一度同期を確立したら、当該同期のタイミングを維持し、ハッシュが一致していることのみを確認する
- ハッシュが不一致になった場合には伝送のジッタが想定されるため、現タイミングの状態から前後に探索領域を拡げつつ、ハッシュの一致するデータを探索するという処理により、唯一性が得られなかった場合の同期外れを防止しつつ、ジッタ時の自動調相を確立している。

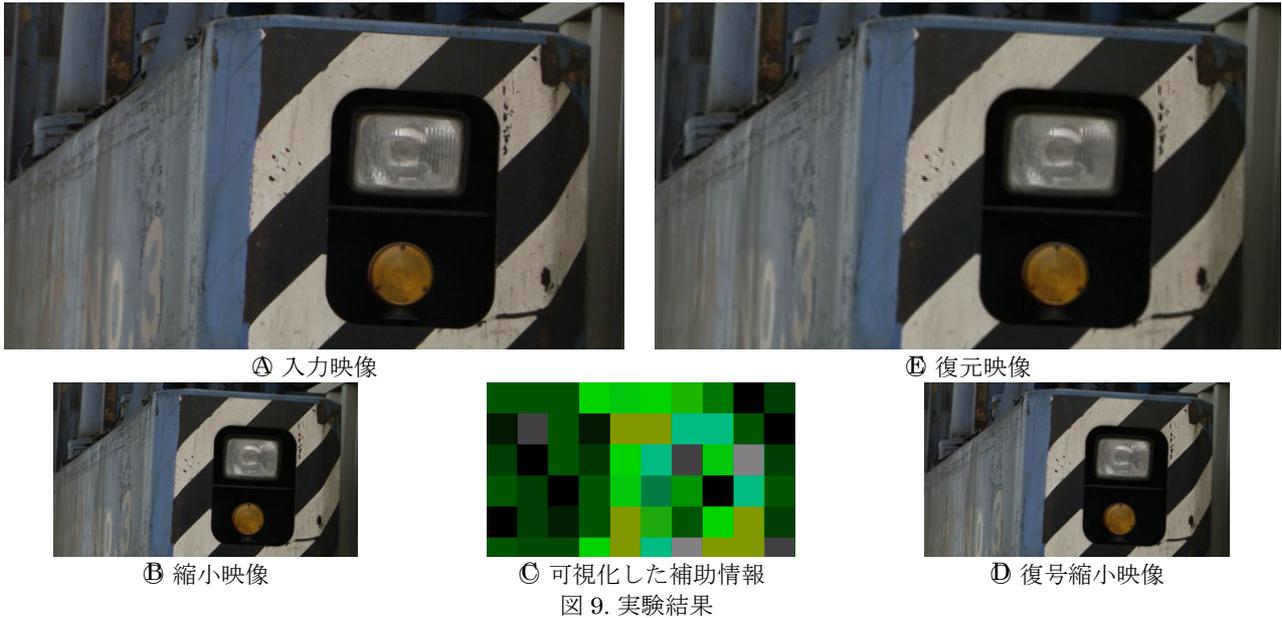


図 9. 実験結果

ハッシュ同期のほか、画面左上の 32×32 領域に白/黒のマーカを 10 秒周期で交互表示する同期手法（マーカ同期）も実装した。こちらは、送信側における入力映像と局部復号映像の同期のため事前調整に使用するほか、ハッシュ同期方式に代えて受信側における映像ストリームと補助情報ストリームの同期のバックアップとしても利用可能である。



図 7. 技研公開 2015 での実演



図 8. リアルタイム空間解像度・階調変換装置の内部

3. 予備的な実証実験

今回開発した装置を用いて 4K 映像のリアルタイム符号化の実演を行った。図 7 に 2015 年 5 月に開催された NHK の技研公開 2015 における実演のようすを示す。図中、送信側前処理および受信側後処理と示したものが、今回開発した空間解像度・階調変換装置である。図 8 は装置の内部である。右に垂直に設置された 4 枚の FPGA 基板では超解像復元 0~3 の各処理と図 5 の最適化処理が行われる。

従来型の映像符号化には HEVC/H.265 を用いた。ビットレートの割り当ては、2K 縮小映像に対し 15.5Mbps、補助情報に対し約 0.5Mbps の合計約 16.0Mbps である（補助情報は非圧縮で 735kbps を要するが、ロスレス符号化により 60% 程度にまで圧縮している）。すなわち、現行の 4K 試験放送（約 35Mbps）の約半分の帯域で 4K および 2K の 2 解像度の映像伝送を実現している。

システムの各要所における映像（部分拡大）を図 9 に示す。図中の④~⑥は図 1 の各吹き出しと対応している。

今回は超解像復元 0~3 は、超解像なし、超解像弱、超解像中、および超解像強の 4 種類とした。概して、輝度については④のブロックが明るいほど（色差についてはマゼンタが強いほど）強い超解像が選択されたことを表している。絵柄の平坦な場所は超解像なしまたは超解像弱が選ばれているのに対し、縞模様やライトなどのテクスチャ部には強めの超解像が選択されている。

4. まとめ

超高精細映像のリアルタイム伝送を目指して、従来符号化前後で空間解像度および階調を変換することにより高圧縮時の画質を改善するシステムを開発した。前後処理、ブロック単位での最適化、および映像と補助情報の同期を含めシステムのリアルタイム動作を確認した。現在、前後処理における超解像や階調変換のパラメータ選定、最適化規範における各指標のバランス、画質の客観/主観評価など、システムを最良の状態に運用するための検証を進めている。

参考文献

- [1] T. Misu, Y. Matsuo, S. Iwamura, S. Shishikui: "Real-time Video Coding System for Up to 4K 120P Videos with Spatio-temporal Format Conversion," ICCE 2015, pp.58-61, Jan. 2015.
- [2] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli: "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol.13, no.4, pp.600-612, Apr. 2004.