

画像・映像圧縮符号化における視覚特性の利用

高村 誠之

NTT サイバースペース研究所
〒 239-0847 横須賀市光の丘 1-1 Y-517A

1. はじめに

JPEG, MPEG-2, H.264/AVC などに基づく符号化技術は、画像・映像を元の容量の 1/10~1/1000 にまで小さく圧縮することを可能とする技術である。現在個人や一般家庭にまで普及している、

- 地上波デジタル TV 放送やワンセグ放送
- DVD や Blu-Ray ディスク, ハードディスクレコーダ, デジタルカメラ, 携帯電話での写真・動画の撮影や視聴
- ネットワーク経由での動画共有や番組視聴等の映像コミュニケーションシステムは、この符号化技術が根幹となり支えている。

画像・映像符号化方式は、原信号が復号信号に完全一致する「可逆符号化」と、一致しない「非可逆符号化」の二種類に大別されるが、本稿でいう符号化は、後者の「非可逆符号化」を指すこととする。画像・映像は様々なメディアを通して送受されるが、最終端末は多くの場合

人間であり、そのため非可逆符号化方式においては、画質の「主観的な評価」が極めて重要である。そこで、SNR のような「客観品質」のみならず、主観品質を高めるための様々な視覚特性利用が検討されている。

図 1 に、ある画像 (a) に対し等しい電力 (mean square error, MSE) の歪が重畳した 2 画像 (b), (c) の例を示す (文献 1 より引用)。明らかに (b) より (c) の歪の方が大きいように我々の目には知覚される。非可逆符号化の場合、当然ながら復号画像には歪が重畳する。従って、符号化の品質改善は「符号化歪電力一定のもと、如何に主観品質を高めるか」、換言すれば「如何に符号化歪を人間の目に知覚されにくい信号空間へ掃き寄せるか」にかかっている。

本稿では、画像・映像符号化の基本的な流れを踏まえた後、視覚に関する過去の知見を基にどのような「歪の掃き寄せ」が実際に行われているか、実例を示す。

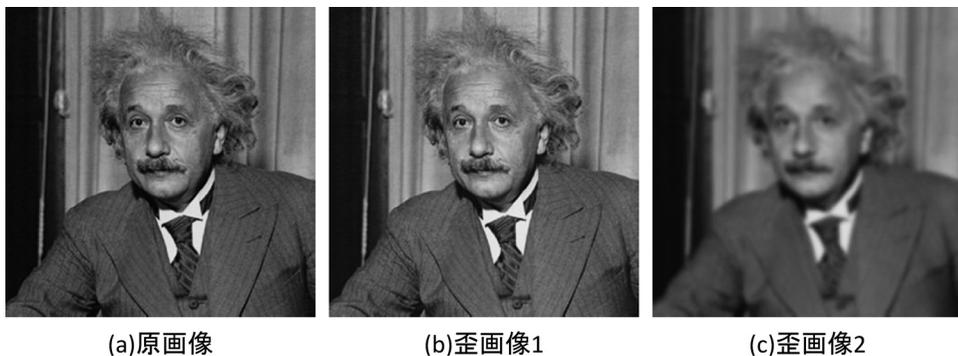


図 1 同じ MSE での主観画質の差. 原画像 (a) に対し歪画像 (b), (c) はともに MSE=144 である.

著者は視覚研究の専門家でないため用語や引用は必ずしも正確・適切でない場合があるかもしれないがその点は予めご容赦を賜りたい。

2. 画像・映像の符号化手順

画像・映像の符号化は基本的に以下の4ステップを入力に順に施すことで行われる(図2参照)。

1. 予測—空間的予測(周辺の復号済み画素からの画素値予測)あるいは時間的予測(過去・未来の復号済みフレームからの動き補償予測)。画像信号のエントロピを低減する。
2. 変換—離散コサイン変換(DCT)あるいはそれに類似した直交変換。信号の頻度分布の偏りを強め、さらにエントロピを低減する。
3. 量子化—変換係数を離散化し整数番号をふる。逆量子化後、元の係数値は得られないため、符号化処理が非可逆となる根本の原因となっている。量子化により大半の変換係数が0となるため、また量子化係数はジグザグスキャンによる一次元化を経て零係数のつながり(run)と非零係数の値(level)他を組み合わせる情報源統合が施されるため、大幅にエントロピが低減される。また量子化幅の増減による発生符号量の調整も可能となる。
4. エントロピ符号化—上述の一連の処理の結果得られた離散信号にハフマン符号・算術符号などの情報源符号化を施し、冗長性の少ないビット列を得る。

符号化の前段階の処理で行われる視覚特性利用を第3章で紹介したのち、「歪の掃き寄せ」に

強くかかっている上記2. 変換, 3. 量子化と視覚特性の関係について第4, 5章で説明したい。

3. 色覚感度—輝度色差空間における空間間引き

人の色覚は、輝度には敏感で色差には比較的鈍感である、という事実はよく知られている。RGB各8bitの信号を輝度色差空間へ写像する一例を以下に示す：

$$Y=0.257R+0.504G+0.098B+16$$

$$Cb=-0.148R-0.291G+0.439B+128$$

$$Cr=0.439R-0.368G-0.071B+128$$

Yが輝度、Cb, Crが色差である(Cbが青—黄軸, Crが赤—緑軸に相当する)。この写像は逆写像を持つため、YCbCr値をRGB値へ戻すことができる。

ある画像を、RGBの3プレーンおよびYCbCrの3プレーンに分解したものをそれぞれ図3および図4に示す。RGBと異なり、YにはCb, Crよりも多くの空間的情報が含まれていることがわかる。そこで符号化前に

1. RGB→YCbCr空間への写像
2. Cb, Cr信号の空間的間引き

という処理を施すことが一般に行われている。

代表的な色差間引きを図5に示す。広く用いられているのはCb, Crをそれぞれ縦横1/2に間引く「4:2:0フォーマット」である。実際、RGB→4:2:0→RGB(すべて各段階8bit)と変換しても、原画と殆ど見分けがつかない(しかしながら、圧縮率をかなり高くすると4:2:0では色にじみがより強くなる)。4:2:0フォーマットの場合、サンプル数比はRGBあるいは4:4:4フォーマットの場合の半分となり、圧縮に有利に働く。

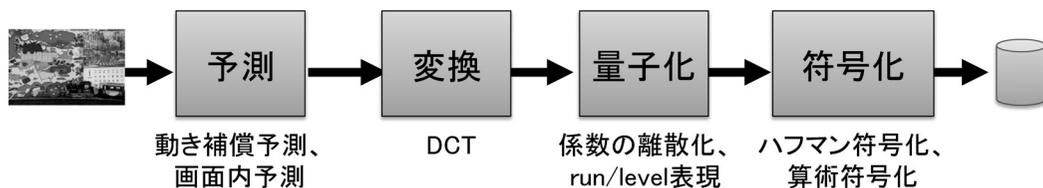


図2 画像・映像符号化の処理の流れ。



原画像



R成分



G成分



B成分

図3 カラー画像とそのRGB成分.



Y成分



Cb成分



Cr成分

図4 図3の画像のYCbCr成分.

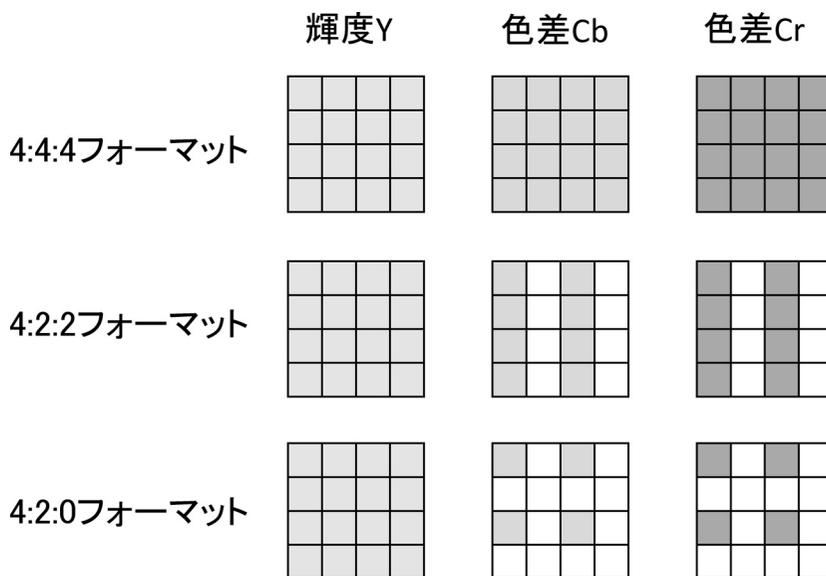


図5 色差信号の空間間引き. 4:4:4比サンプル数は4:2:2フォーマットが2/3, 4:2:0フォーマットが1/2.



図6 マスキング効果の実例. 原画像 (a) およびそれに MSE=32.3 の加法的白色ガウス雑音を加えた雑音重畳画像 (b).

4. 空間的な画質変化への感度

4.1 マスキング効果—アクティビティに基づく量子化粗密調整

図6にマスキング効果の実例を示す. 原画像が同図 (a) で, それに MSE=32.3 の加法的白色ガウス雑音を加えた画像が同図 (b) である. 帽子飾りの部分はほとんど雑音が知覚できないが, 肩や背景はうっすらと雑音が知覚できる.

MPEG-2 の Test Model 5²⁾ が用いる以下の 3

段階の符号量制御は上記知見に呼応したものであり, 同一符号量において SNR は 1dB ほど低下するが, 主観品質は向上する.

1. ピクチャタイプに応じた, 一フレームへの符号量割り当て
 2. 符号化対象マクロブロックの量子化幅を, それまでの発生符号量に鑑み設定
 3. 該マクロブロックの「アクティビティ」に応じた量子化幅の局所増減
3. の「アクティビティ」は, 下記で act として

求められる。本処理では、2. で求められた量子化幅を下記で求められる N_{act} 値に比例させ増減させる。

$$act = 1 + \min(var_0, var_1, var_2, var_3)$$

$$N_{act} = \frac{2act + avg_act}{act + 2avg_act}$$

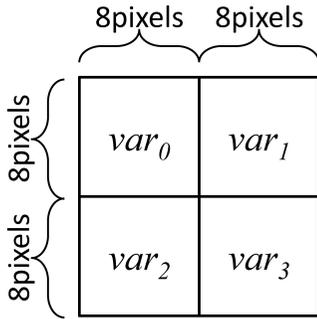


図7 16×16のマクロブロック内で分散を求める小ブロック。

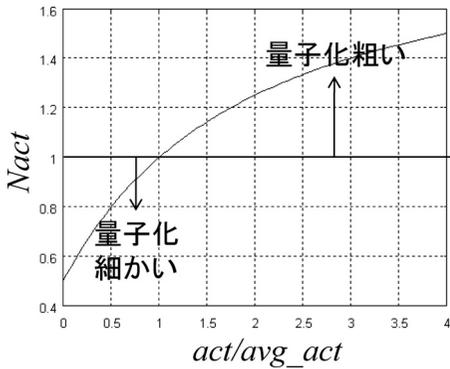


図8 act と N_{act} の関係。

ここで var_n は 8×8 のサブブロック内の画素値分散 (図7参照), avg_act は以前に符号化した画面でのアクティビティ act の平均である。 act と N_{act} の関係を図8に示す。 N_{act} は 0.5 から 2 までの値をとり、アクティビティ act が大きいほど量子化が粗く、小さいほど細くなるよう制御される。図6(a)における各マクロブロックの N_{act} 値を可視化したものを図9に示す。画像 baboon において、図9と同様に N_{act} 値を可視化したものを図10(a) (b) に示す。

図6(b) で直感的に感じる歪許容度とよく合致する指標が N_{act} により定量化できていることがわかる。

4.2 コントラスト感度関数—量子化マトリクス

図11にコントラスト感度関数 (CSF)³⁾を可視

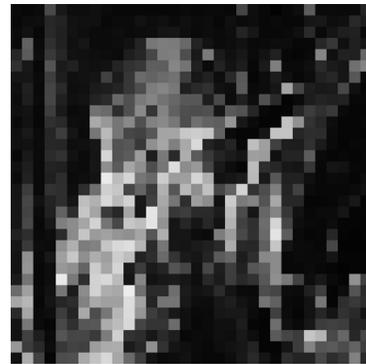
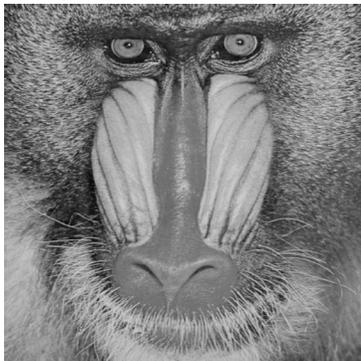
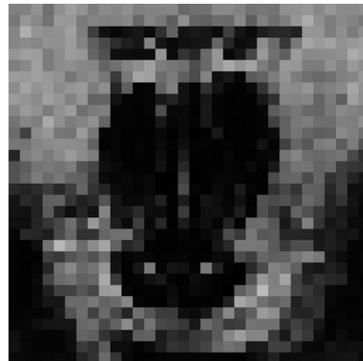


図9 図6(a)における各マクロブロックの N_{act} 値を可視化したもの。 N_{act} が大きいほど白く、小さいほど黒い。



(a)原画像(baboon)



(b) N_{act} 値画像

図10 画像 baboon (a) において、図9と同様にマクロブロックの N_{act} 値を可視化したもの (b).

化するための評価画像を示す。左上や右上の縮模様が知覚されにくいことがわかる。本稿ではこれが変換後の信号に対して用いられている例を紹介する。

図 12(a) は H.264/AVC の参照ソフトウェア⁴⁾ がデフォルトで用いている量子化幅修正係数を周波数ごとに並べたもので、一般に「量子化マトリクス」と呼ばれるものである。同図 (b) はそれを 3 次元グラフにしたものであるが、低周波成分ほど量子化係数を低めにとり、高周波成分ほど量子化係数が拡大されることを示している。図 11 では超低周波は逆に感度が下がっているが図 12 ではそのようになっていない（直流

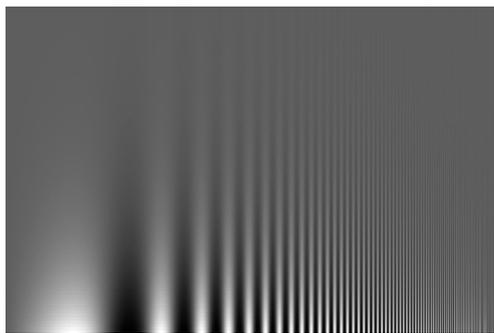


図 11 コントラスト感度関数を可視化するための評価画像。

が最も歪が小さくなるようになってい）この理由としては、この変換のような 8×8 程度のブロックサイズでは、直流の変換係数は波長 8 以上の交流を表している可能性もあるため、と説明できる。

CSF のより積極的な量子化への応用例としては例えば文献 5 などを参照されたい。

5. 時間的な画質変化への感度

5.1 時間鋭鈍交番符号化

図 13 に示す符号化構造は一般によく用いられているものの一つである。通常 B フレームの量子化幅は、I/P フレームのそれに比べ大きめに設定される。B フレームが他から参照されない場合はその重要度が参照されるフレームより低いためである（参照される B フレームの場合はこの限りではない）。必然的に B フレームの符号化歪は他よりも大きいものとなるが、視覚的にはそれほど大きな劣化として知覚されない。このような歪の時間変動は、映像を後で切り貼りする編集用途には向かないが、それ以外の用途であれば、（インタレース 60Hz やプログレッシブ 30Hz 程度のフレームレートで）2-3 フレーム程度までこのような低画質画像を挿入し、符号量を削減することが積極的に行われている。

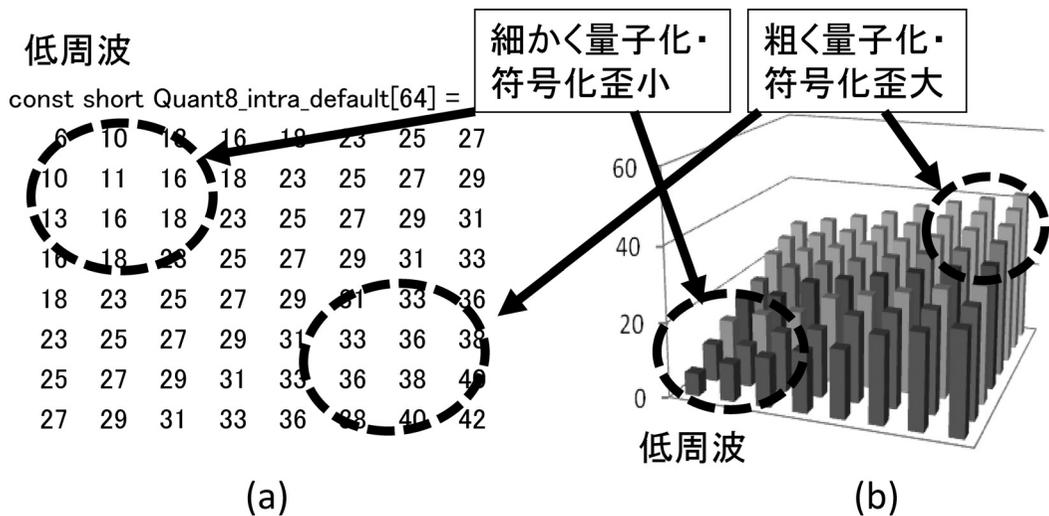


図 12 周波数空間の「歪き寄せ」の例。H.264/AVC 参照ソフトウェア JM のデフォルト量子化マトリクス (a) およびそれを可視化したもの (b)。

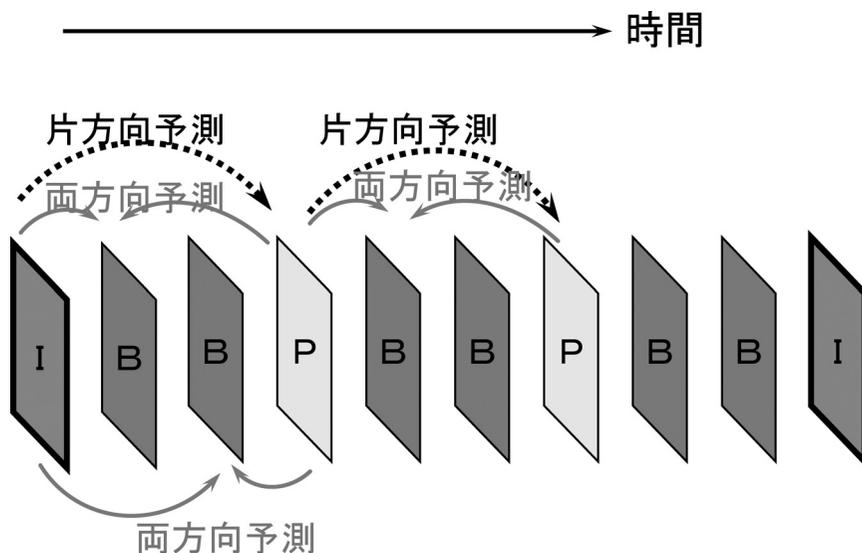


図 13 符号化構造の一例. I: 画面内符号化フレーム, P: 片方向予測フレーム, B: 両方向予測フレーム

5.2 コントラスト感度関数の時間変化

速く動く物体は遅い物体より知覚が困難になるということは日常よく経験される。つまり、空間周波数に対するコントラスト感度は、時間軸方向の変化量に依存し、例えば時間軸方向の大きな変化（高速なカメラパン・チルト、シーンチェンジ等）が発生したフレームでは、そのフレームの画質劣化に対する感度は低下する。このため、時間軸方向の大きな変化が発生したフレームに対しては、CSF を小さめに見積もるよう制御することで、さらなる符号量削減が期待できる。坂東らは物体速度の大小を視距離の遠近に置き換えた「時空間視覚感度関数」を定式化⁶⁾、主観品質を保ったまま符号量を 5% 以上削減することに成功している。

6. おわりに

画像・映像符号化における視覚特性を利用した、色空間、時間、周波数空間、画素空間における「歪の掃き寄せ」事例を紹介した。

符号化研究と視覚研究は、一見互いの研究領域が異なるように思えるが実は評価者が人という点で強く繋がっている。視覚の研究成果を符号化へ応用できないか積極的に考えることや、

さらに視覚研究の視点から符号化技術を再検討することも今後重要になるものと思われる。具体的には、まだ符号化研究者が知らない「未知の不可視空間」を発見し、符号化の枠組みにうまく組み入れる、という serendipity を期待したい。

文 献

- 1) <http://www.ece.uwaterloo.ca/~z70wang/research/ssim/>
- 2) ISO/IEC JTC 1/SC 29/WG 11 N0400 “TEST MODEL 5”.
- 3) F. W. Campbell and J. G. Robson: Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, **197**, 551–566, 1968.
- 4) Joint Video Team: JM Reference Software. available at <http://bs.hhi.de/~suehring/tml/>
- 5) C. Jianwen, Z. Jianhua and H. Yun: Macroblock-level adaptive frequency weighting for perceptual video coding. *IEEE Transactions on Consumer Electronics*, **53**(2), 775–781, 2007.
- 6) 坂東幸浩, 早瀬和也, 高村誠之, 上倉一人, 八島由幸: ブロック歪みを考慮した時空間コントラスト感度特性に基づく H.264/AVC 符号化器設計. *FIT2008, RI-004*, 13–16, 2008.