

高能率符号化動画像の画質に関する基礎的研究

半谷 精一郎

東京理科大学 工学部

電気工学科 助教授

1. はじめに

高能率符号化画像の画質評価にSNR (Signal to Noise Ratio) を用いることが不適切であることは一般によく知られている〔1〕。これは、符号化時に生じる雑音が、輝度が大きく変化するエッジ付近に集中したり、ブロック状に分布して、雑音量が同じでも主観的にはまったく異なって感知されるからである。そのため、人間の視覚の空間周波数特性を正弦波グレーティングを用いて測定し、雑音量を計算する際にこのことを考慮する方法〔2〕や、実画像に加えた空間周波数選択性雑音に対するMOS (Mean Opinion Score) をもとに導出した重み付け関数を利用するWSNR (Weighted SNR)〔3〕などが提案されている。

このうち、後者のWSNRに関しては、さらに原画の局所的なフラクタル次元を利用して改良したAWSNR (Advanced WSNR〔4〕)がMOSと0.93の相関を与えており、DCT (Discrete Cosine Transform)〔5〕のような変換符号化やVQ-PE (Vector Quantization of Prediction Error)〔6〕のような予測符号化によって高能率符号化された画像の80%のMOSを±0.5の確度で推定できることを明らかにしてきた〔7〕。

そこで、本研究では動画像符号化の国際標準規格であるH.261〔8〕をもとに符号化した動画像の画質をAWSNRを用いて評価し、フレームの進行とともにMOSがどのように変化するかを推定するとともに、実際に主観評価を行った結果と比較する。

2. フラクタル次元を利用するAWSNRについて

WSNRは画像の空間周波数領域における雑音成分に視覚の空間周波数特性に相当する重み付け関数を乗じ、人間には認知されにくい高周波雑音を評価しないようにした客観評価尺度である。静止画像を数種類の符号化方式により符号化し、その画質をWSNRにより評価すると、WSNRのほうがSNRよりMOSを推定しやすいことが判明している。ただWSNRは画像の種類によってはMOSを推定しにくいということもわかっており、何らかの対応が望まれていた。

そこで、フラクタル次元を用いて画像の局所的な複雑さを考慮し、複雑な絵柄の上の雑音は検知しにくいという人間の特性をWSNRと組み合わせて改善してAWSNRを提案してきた。すなわち、多数の様々なフラクタル次元 (F_{dim}) をもったテクスチャ画像を用意し、これに様々な大きさのランダム雑音を加えた画像の主観評価実験を行なうことにより、図1のようなMOS - WSNR - F_{dim} の関係を求めた上で〔4〕、次のような手順によりAWSNRを定めることにした。

- (1) フラクタル次元の求めやすさと局所的な特徴を反映させるために、原画像を32画素×32画素の大きさのブロックに分割する。
- (2) 原画像と符号化画像とから各ブロックのフラクタル次元とWSNRを求める。

- (3) WSNRとフラクタル次元から各ブロックのMOSを推定し、そのMOSを与える $F_{dim}=2$ の時の $WSNR_2$ から、 $WSNR_2/WSNR$ で与えられる雑音のマスク係数を求め、各ブロックの雑音を軽減する。

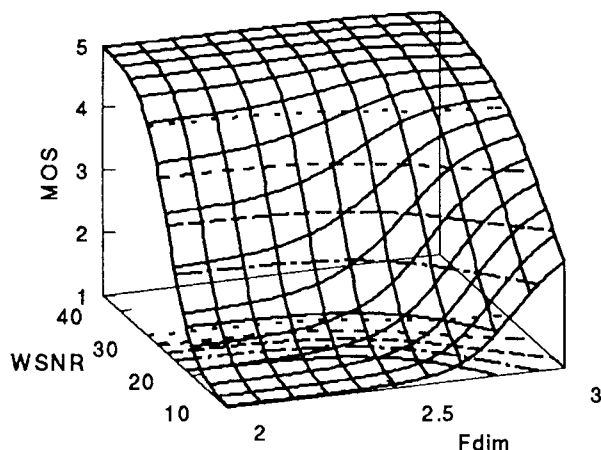


図1 12枚のテキスチャから求めたMOS-WSNR-F_{dim}の関係〔4〕

- (4) 全ブロックの雑音をこうして軽減し、画像全体の平均から求めたSNRをその画像のAWSNRとする。図2に4種類の画像(ITE-HAIRBAND, ITE-CHURCH, SIDBA-MILKDROP, CCITT-BARBARA)に5種類の符号化方式(DCT, A-DCT, DCT-VQ, A-DCT-VQ, VQ-PE2)およびランダム雑音を付加した合計200枚の画像に対するMOS-WSNR特性とMOS-AWSNR特性を示す。図中の一点鎖線は、測定値との2乗誤差を最小にする次式のようなロジスティック曲線〔9〕を与えるもので、

$$MOS = \frac{4}{1 + \exp\{-\alpha(WSNR - \beta)\}} + 1 \quad (1)$$

MOS-WSNR特性においては $\alpha=0.28$ 、 $\beta=25.09$ 、MOS-AWSNRにおいては $\alpha=0.348$ 、 $\beta=27.80$ となった。また、これらの結果からWSNRもしくはAWSNRからMOSを ± 0.5 の確度で推定するとすればWSNRでは56%の画像しか該当するものがないが、AWSNRでは80%の画像に適用できることも判明した〔7〕。

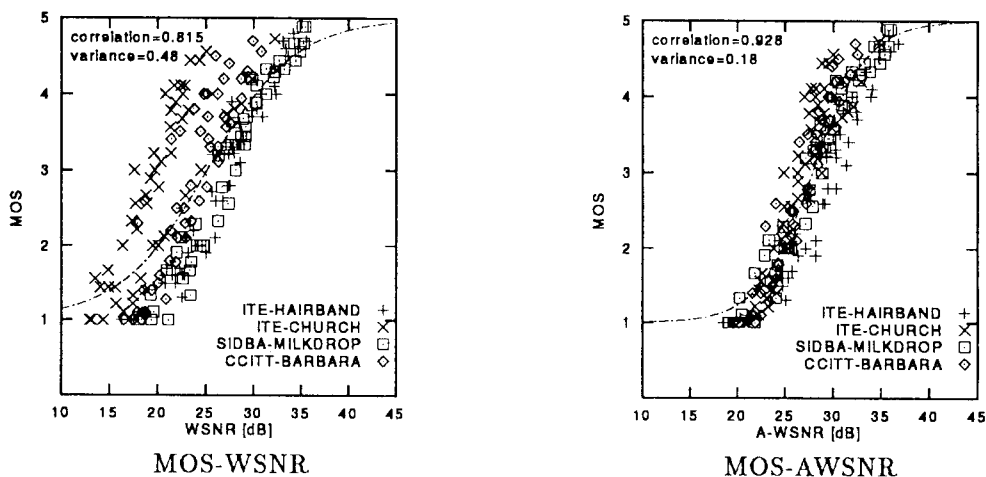


図2 MOS-WSNR特性とフラクタル次元を考慮して改善したMOS-AWSNR特性

3. 動画像の動的MOSの推定

符号化された動画像の画質を主観的に評価することは、国際規格を定める際などは必ず行なわれるようであるが、その客観評価尺度となるとSNR以外は見当たらない。しかし、SNRが高効率符号化動画像の適切な評価尺度となりうるかといえばやはり否定的である。そこで、視覚の時間周波数特性などを考慮せずに上記のAWSNRを胸上画像であるMiss America (150フレーム)の各フレーム毎に求め、その値からMOSを推定することを試みた。

その際の符号化方式としては、H.261で規定されているMC+DCT方式(以降、これを方式1と呼ぶ)と、リフレッシュ直後はMC方式により符号化し、途中からMC+DCT方式に切替える方式(10)(以降、これを方式2と呼ぶ)の2つを扱った。この時の、伝送符号量は256kbps、フレームレート30frame/secとし、DCTを施すブロックサイズ: 8×8 画素、動きベクトル探索範囲: ± 7 画素、リフレッシュフレーム: 61フレーム目とした。

図3に、方式1ならびに方式2により符号化した動画像のMOSをAWSNRをもとに推定した結果を示す。また、図4に方式1ならびに方式2のSNRがフレームの進行とともにどのように変化するかを示す。どちらの方式でも第2フレームまでは、伝送符号量の制約から推定MOS、SNRとも悪く、かなりの画質劣化が生じていることがわかる。一例として、第2フレームの原画像と符号化画像(方式1および2によるもの)を図5に示す。この図から、主観評価した場合には、ほとんどの評価者のMOSが1もしくは2となることが予想され、図3の推定MOSとよく一致すると思われる。

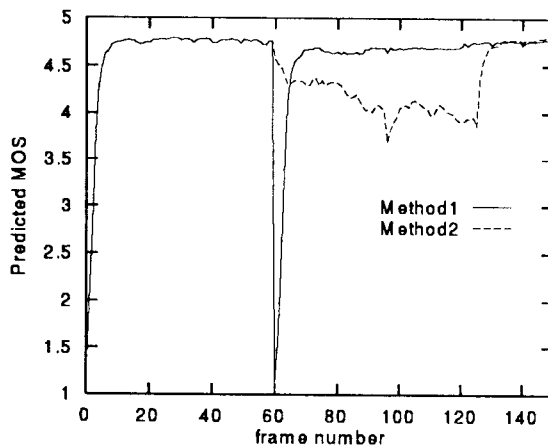


図3 AWSNRから求められた
動画像の推定MOS

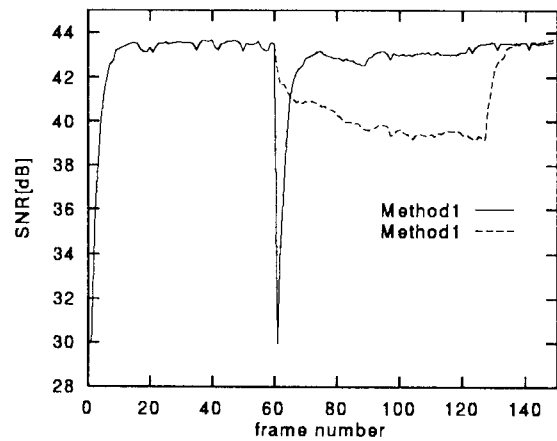


図4 動画像のSNR

また、第3フレーム以降についても実際に主観的に画像を観察したところ、第60フレームまでは両方式の画像にはあまり劣化がみられず、図3および図4の結果はいずれも主観評価値を反映したものとなった。

ただし、第61フレームのリフレッシュ後、第64フレームまでは、方式1の劣化が主観的にも顕著にわかった。一例として、図6に第63フレームの原画像と符号化画像(方式1によるもの)を示す。これは方式1がリフレッシュ時に低ビットレートでINTRA符号化を行ない、それ以降はMCで得た動きベクトルを用いながら予測誤差も符号化するため、伝送容量の上限である



図5 第3フレーム画像（左から、原画像、方式1による画像、方式2による画像）

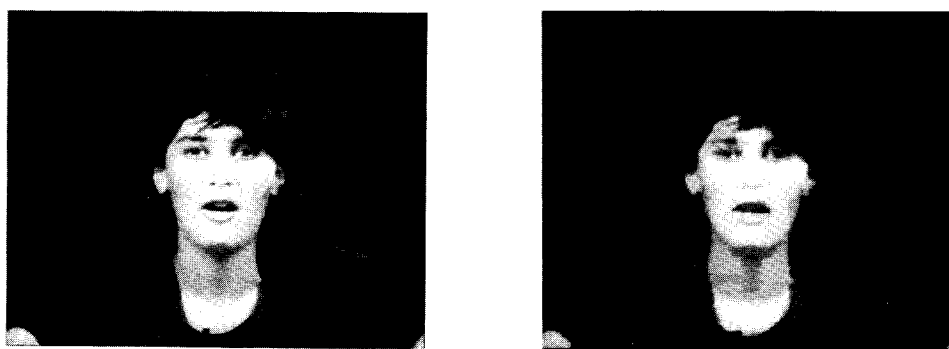


図6 第63フレームの画像（左か、原画像、方式1による画像）

256kbps以上の情報が発生しているためである。このことは、図3の推定MOSが1.8になり、その後徐々に回復しているが、64フレーム辺りまでMOS < 4であることと一致している。一方、図4では第63フレーム以降のSNRが37dB以上となり、SNRが37dB以上だと画質の劣化が認められないという一般的な見解からは画質の低下は読みとれない。このように方式1による符号化画像に関しては、SNRよりもAWSNRをもとにして求めた推定MOSの方が主観評価値に近いことがわかった。

これに対し、方式2では、リフレッシュ前の画質は、実際は劣化がわかる場合は少なくMOSは5に近いはずである。実際、図3より、リフレッシュ前の推定MOSは4.7前後の値をとっており、MOSをよく反映しているといえる。また、図7で示したようにリフレッシュ後の画像中



図7 第97フレームの画像（左から、原画像、方式2による画像）

第97フレーム辺りではほとんどの人が劣化に気付いており、3.9という推定MOSは主観評価結果をよく反映していることがわかった。一方、SNRは39.5dB~43dBの変化しか示しておらず、一般的な見地からはMOSが4以下であるというのは難しく、本研究で提案したAWSNRがSNRに比べ、より我々が感じる劣化量に対応していることを裏付けている。

4. まとめ

フラクタル次元を用いて画像の局所的な複雑さを考慮しつつ符号化画像の画質の客観評価を与えるAWSNRによれば80%までの高能率符号化画像のMOS値を±0.5で推定できることを示し、2つの動画像符号化方式の特性を与えることへ応用してみた。

その結果、従来のSNRよりもMOSを直後推定でき、その値も主観的なものに比較的良好一致することがわかった。

今後は、時間方向の視覚の特性を考慮した評価尺度を確立し、より精度の高いMOSの推定を行えるようにする必要があると考える。

謝辞 本研究を遂行するにあたり、(財)高柳記念電子科学技術振興財団より多大のご援助を賜りました。ここに、理事長はじめ関係の皆様へ深謝致しますとともに、謹んで御礼申し上げます。

参考文献

- [1] 宮川, “テレビジョン画像の評価技術”, コロナ社, 1986
- [2] 宮原, 小谷, 西村, “画像の客観評価尺度—画質の劣化要因の貢献度”, 信学技報, IE87-112, pp.1-8, 1987
- [3] 半谷, 和田, 宮内, “静止画の画質を評価するための空間周波数領域上の重み付け関数の実験的導出”, テレビ誌, vol.46, No.3, pp.295-299, 1992
- [4] Hangai, Suzuki and Miyauchi, “Advanced WSNR for Coded Monochrome Picture Evaluation using Fractal Dimension”, Proc. of Picture Coding Symposium, Sacramento, CA, pp92-95, Sep. 1994
- [5] 相澤, 原島, 宮川, “画像信号適応的離散コサイン変換ベクトル量子化”, 信学論(B), J69-B, No.3, pp.228-236, 1986
- [6] 伊東, 内藤, 佐藤, 宇都宮, “画像の予測誤差信号ベクトル量子化”, 信学論(B), J70-B, No.3, 1987
- [7] 白井, 半谷, 宮内, “画像のフラクタル次元を考慮したWSNRによる動画像の画質の検討”, 画像符号化シンポジウム予稿集, 2.3, Oct. 1994
- [8] 原島, “画像情報圧縮”, オーム社, 1991
- [9] 半谷, 石川, 宮内, “静止画の客観評価尺度について(その5)”, 画像符号化シンポジウム予稿集, pp.7-8, Oct. 1990
- [10] 鎌田, 半谷, 宮内, “低遅延動画像符号化を実現する効果的なリフレッシュ方式の提案”, 画像符号化シンポジウム予稿集, 11.3, Oct. 1994