

2-6 人の主観画質を定量化する研究の難しさと楽しさ

Difficulties and Accomplishment in the Study of Objective Evaluation Measure Considering Human Visual System

半谷精一郎

1. はじめに

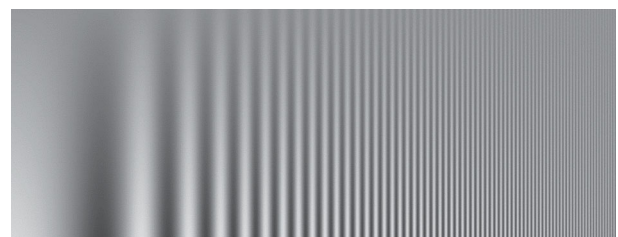
テレビジョン放送が開始されて以来、人類は画像として遠くの情景を見たり、過去の情景を見たり、あるいは記録できるようになった。我が国のアナログ放送の主役であった NTSC 方式はそれまでのモノクロテレビジョンの国際標準である M 方式にカラー信号が多重化されたもので、走査線は 525 本、インタレース走査を行って毎秒 30 フレームで画像が表示されていた。これらの数値は、当然のことであるが‘ひと’の輝度情報に対する時空間周波数特性を考慮したものであった。更に、このときの輝度信号の帯域は 4.2 MHz であったが、色信号の帯域は I 信号が 1.3 MHz、Q 信号が 0.6 MHz しかなかった。これも、‘ひと’の色情報に対する特性によるものである⁽¹⁾。やがて情報通信技術の進歩により、信号はアナログからデジタルに移行する。静止画像の符号化の標準である JPEG が 1992 年に定まり、高圧縮率でも画質劣化が起きにくい JPEG2000 が 2000 年にリリースされた⁽²⁾。一方、動画の初期の符号化の標準である MPEG-1 は 1993 年に、それまでの各種放送方式と互換性がある MPEG-2 は 1994 年にリリースされ、その後、2K、4K という高精細テレビジョン放送を可能にできた⁽³⁾。

こうした符号化の標準を定める上で、最終的な受け手である‘ひと’の主観画質は重要な役割を担っている。例えば、偽輪郭やブロック雑音という、符号化画像に生じる特有の妨害は主観画質を著しく劣化させるが、原画像との差で求められる SN 比だけではうまく説明できな

かった⁽⁴⁾。

そこで、画像の符号化に関する研究が盛んに行われていた 1990 年代の前半から‘ひと’の主観を反映するような客観評価尺度を作ることを目指して研究を行ってきた^{(5)~(8)}。また、2004 年には本会の中にイメージ・メディア・クオリティ時限研究専門委員会（現イメージ・メディア・クオリティ研究専門委員会）の設立をお認め頂き、国内外の研究者の意見交換の場を提供できるようになった。

以下では、どのようなきっかけで画質の研究に取り組んできたか、また、派生する研究を行ってきたかをお伝



(a) 正弦波グレーティング



(b) 視感度

図 1 正弦波グレーティングによる空間周波数特性の求め方 (a)の視距離を変えていくと‘しま’模様が見える部分が(b)の塗りつぶし領域のようになって、‘ひと’の視感度(白線)を求めることができる。

半谷精一郎 正員：シニア会員 東京理科大学工学部電気工学科
E-mail hangai@ee.kagu.tus.ac.jp
Seiichiro HANGAI, Senior Member (Department of Electrical Engineering, Tokyo University of Science, Tokyo 125-8585 Japan).
電子情報通信学会誌 Vol.100 No.6 pp.461-467 2017年6月
©電子情報通信学会 2017

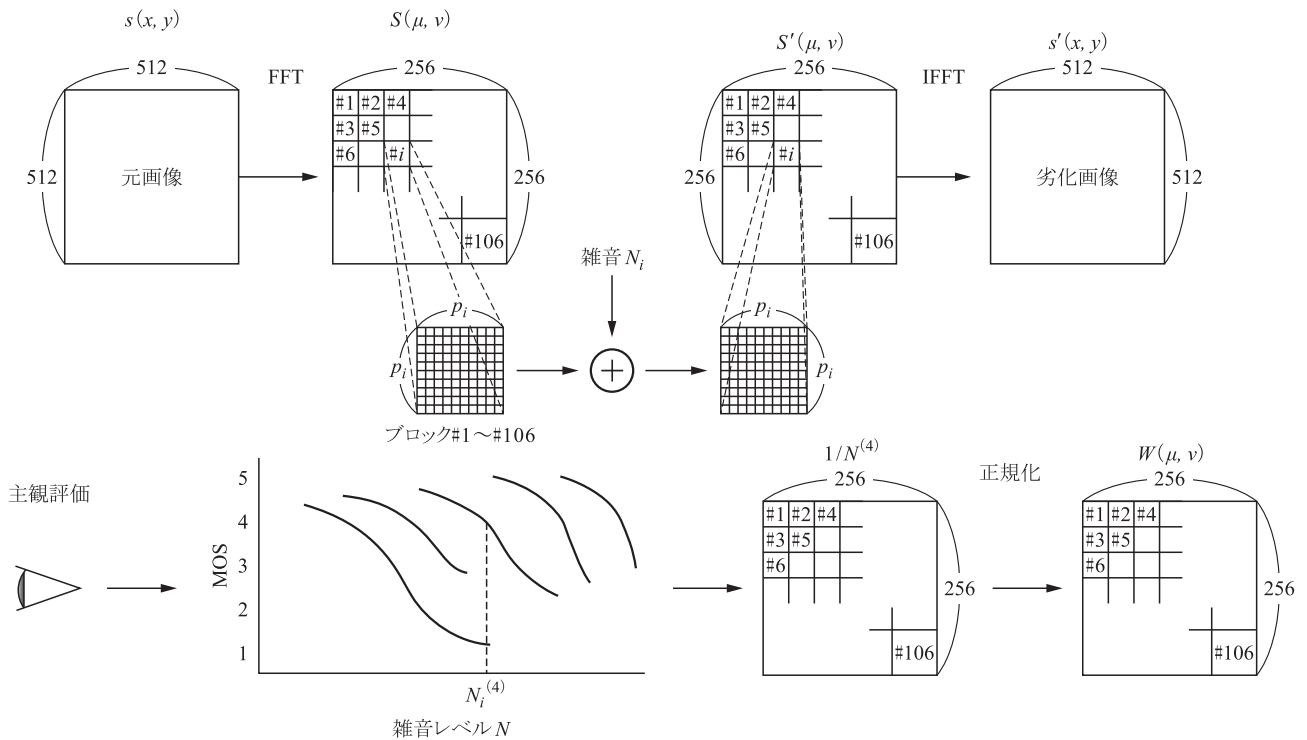


図2 重み付け関数導出過程⁽¹⁰⁾ 106個の空間周波数領域ごとに様々な強度の雑音を加え、MOS=4となる雑音レベルから重み付け関数を求めた。

えし、次世代の研究者へのエールとしたい。

2. ‘ひと’の視覚特性を研究するきっかけ

‘ひと’が知覚する画像の劣化には、ぼけ、ランダム雑音、ブロックひずみ、輪郭の不連続性に関するひずみなど様々なものがあるが⁽⁴⁾、こうした劣化に伴う光刺激があるしきい値を越えたときに認知されることから、多くの場合、50%以上の人認知するかどうか重要となる⁽⁹⁾。例えば、図1(a)のような横方向に空間周波数が増える正弦波グレーティング（縦方向は変調度が増える）を一定の距離で見るときに50%以上の人認知を感じる領域を塗りつぶすと、図1(b)のような境界線が現われる。この境界線が、それぞれの空間周波数に対する‘ひと’の視感度を与える。

一般に2.5~4 cpd（cpdはcycle per degreeで視角1°当りに入る‘しま’の数）が視感度が最も高くなるとされていたが、ブロック符号化との親和性も含め空間周波数特性が等方的であるかどうか検討したいと考え、まずは視覚の空間周波数特性を実験により求めることとした。

2.1 空間周波数領域上の重み付け関数の実験的導出⁽¹⁰⁾

画像のひずみを空間領域で雑音電力として評価すると主観と全く一致しないことが分かっていたので、次式の

ような空間周波数領域で重み付けしたWSN比（WSNR: Weighted Signal to Noise Ratio）を定義し、重み付け関数 $W(\mu, \nu)$ を実験により求めることから研究をスタートさせた。

$$WSNR = 10 \log \frac{255^2}{\sum_{\mu=0}^{255} \sum_{\nu=0}^{255} \{S(\mu, \nu) - S'(\mu, \nu)\}^2 \cdot W(\mu, \nu)} \quad (1)$$

ここで、 $S(\mu, \nu)$ は劣化前の原画像の空間周波数成分、 $S'(\mu, \nu)$ は劣化後の空間周波数成分を与える。空間周波数上で評価するための重み付け関数としては、上述のような正弦波パターンに対する‘ひと’の視覚特性⁽¹¹⁾や、網膜の数学モデルを基に考案された周波数特性⁽¹²⁾などが考案されていた。しかし、ごく一般的な画像を基に求めた重み付け関数の方が実際的な評価値を与えるものと考え、標準画像「肌色チャート」を基に求めることにした。

図2に重み付け関数導出のために考えた処理の流れ図を示す。処理手順は以下のとおりである。

- (1) 標準画像 $s(x, y)$ を二次元フーリエ変換することにより、空間周波数成分 $S(\mu, \nu)$ を求める。
- (2) 空間周波数領域を106個のブロックに分割し、各ブロックに関して、以下の(3)~(5)を繰

り返す。

- (3) i 番目のブロック ($i=1\sim 106$) の空間周波数成分に、様々なレベルのランダム雑音 (雑音電力 N_i) を加え、二次元逆フーリエ変換を行って複数の劣化画像 $s'(x, y)$ を作成する。
- (4) 複数の $s'(x, y)$ を 20 人により 5 段階評価し、MOS (Mean Opinion Score) を求める。
- (5) MOS が 4 (MOS=4 は雑音が検知されるかどうかの主観的なしきい値) となる雑音量 $N_i^{(4)}$ を求める。
- (6) 次式により、ブロックごとの重み付け値 W_i を求める。 (K は定数)

$$W_i = K \frac{1}{N_i^{(4)}} \quad (2)$$

図3にこうして求めた重み付け関数 $W(\mu, \nu)$ の立体表現を示す。ここで μ は水平空間周波数、 ν は垂直空間周波数を与えるが単位はディスプレイ上で表示される‘しま’でTV本である。画像は512画素×512画素なので、256TV本が最高空間周波数となる。視距離を7H (Hは512画素を表示するディスプレイ上の画面高である) とすると、図3のローカルピークである $\mu=\nu=24$ TV本の位置は3cpdに相当することが判明した。この図から分かるように、視感度は等方的な特性ではなく、3cpdの斜め成分に対する感度が高く、ここから8cpdまでは等方的であるが、10cpd以上になるとむしろ斜め成分に対する感度が低下することが明らかになった。このように、標準画像を用いて求めた重み付け関数は、従来から明らかになっていた視感度と大局的には符合したが、ローカルピークの形状や高域での感度な

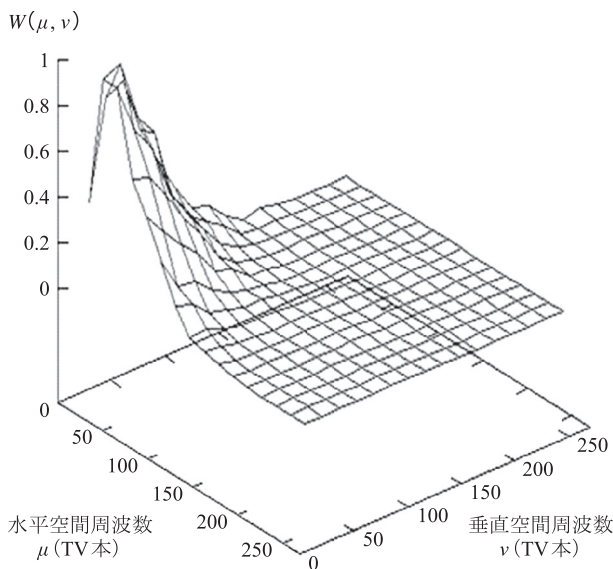


図3 実験的に求められた重み付け関数値 斜め成分に対する感度が高くピークは3cpdに現れたが、視感度は等方的でないことが判明した。

どの点で差異が見られた。

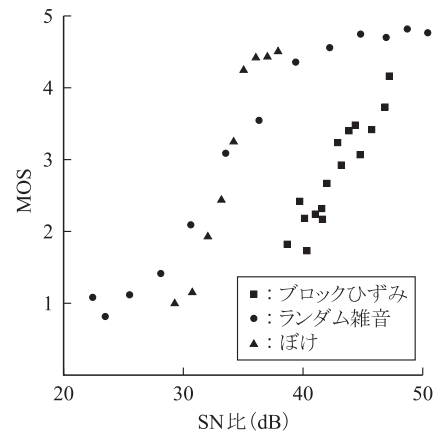
図4(a)は視感度を全く考慮していないMOSとSN比の関係、同図(b)は視感度を考慮したMOSとWSN比の関係を故意にブロックひずみ、ランダム雑音、ぼけの劣化を加えて作った画像を用いて比較した結果である。これらの結果から、視感度を考慮することでMOSとWSN比の相関が高まり、WSN比からMOSを推測できるのではないかと考えられた。

原画像をどのようなものにしても上述の相関は改善したが、改善率は画像によって異なるという結果を得た。

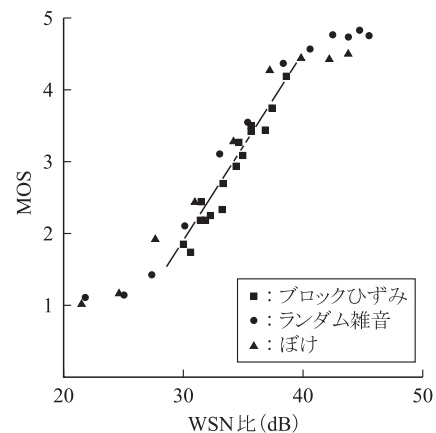
具体的には、同じ位置にある雑音や同じ量の劣化であってもその背後にある画像のテクスチャが異なると図4のような改善が見られないものもあった。つまり、テクスチャを定量化して、それに基づいて雑音や劣化を評価しないと‘ひと’の視覚のマスクング特性を表現できないことが分かった。

2.2 原画像のテクスチャを考慮した客観評価法の提案⁽¹³⁾

雑音や劣化の背後にある画像のテクスチャの複雑さ、



(a) MOS-SN比

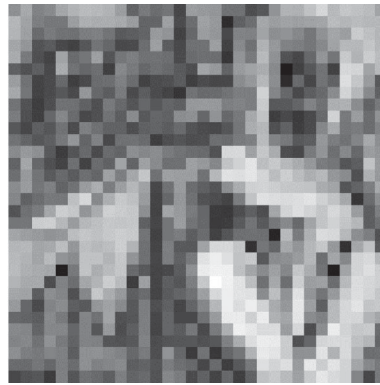


(b) MOS-WSN比

図4 MOSとSN比及びWSN比の関係 よく用いられるSN比を視覚的空間周波数特性を考慮したWSN比にするだけでMOSとの相関が一気に改善された。



(a)原画像



(b)フラクタル次元

図5 原画像から抽出したフラクタル次元
フラクタル次元がよく対応している。

主観的なテクスチャの複雑さとフラク

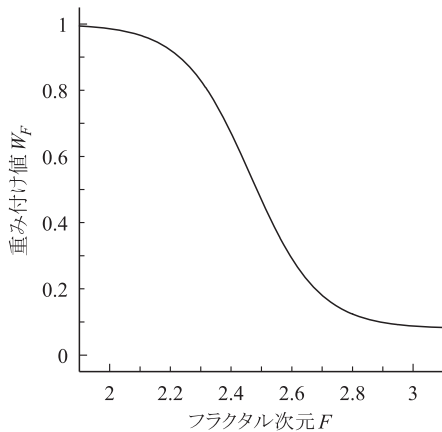


図6 フラクタル次元 F と重み付け値 W_F の関係 雑音・劣化の評価は、平坦な領域 ($F=2$) ではそのままだが、複雑な領域では $1/10$ 以下の評価値となった。

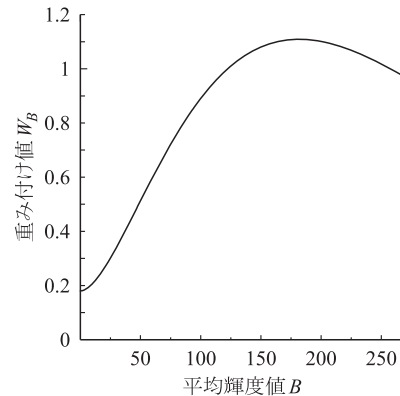


図7 平均輝度値 B と重み付け値 W_B 平均輝度値が暗かったり、明る過ぎた場合にも雑音・劣化が見えにくくなることを確かめた。

局所的明るさ、評価決定領域を考慮した客観評価法を提案し、それによって求めた評価尺度を AWSN 比 (Advanced WSNR) と呼ぶことにした。

画像の局所的な複雑さの定量化は当初エントロピーを使っていたが⁽¹⁴⁾、輝度値のランダムさと複雑さは違うのではないかと考え、いろいろと文献を探していたところ、フラクタル次元という概念に出会った。二次元平面の任意の直線・曲線のフラクタル次元 F は $1 \sim 2$ 、三次元空間の任意の平面や局面のフラクタル次元 F は $2 \sim 3$ と、研究にピッタリの尺度であった。定義式は以下のとおりである。

$$F = 3.0 - \lim_{r \rightarrow 0} \frac{\log V(r)}{\log r} \quad (3)$$

ここで r は立方体の長さ、 $V(r)$ は平面や局面の内側に接する立方体の体積である。

図5は、例として標準画像のフラクタル次元を求めた

ものであるが、雑音や劣化がマスクされそうな次元が高い領域 (白部分) とそうでない領域が検出できていることが分かる。

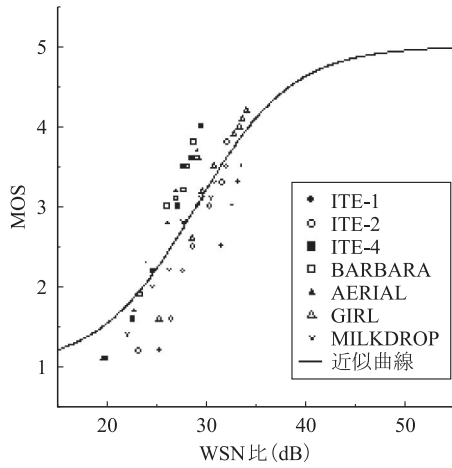
このフラクタル次元 F と実際の主観評価実験から、次式のような重み付け値を導き出すことができた。

$$W_F = 0.08 + \frac{0.92}{1 + \exp\{8.94 \times (F - 2.46)\}} \quad (4)$$

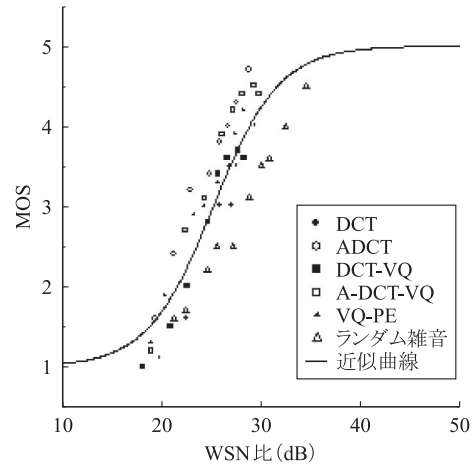
図6に F と W_F の関係を示す。テクスチャが複雑になり、フラクタル次元が高くなるとその上にある劣化や雑音が見えにくくなるという空間領域上の視覚特性をよく表している。

ただし、局所的な輝度値によっては雑音や劣化の見え方も異なることから、これについても実験を行い、図7のような平均輝度値 B と重み付け値 W_B の関係も求めた。

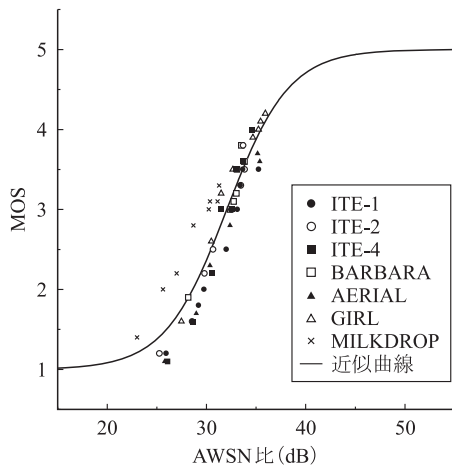
更に、主観評価時間が10秒に制限されると評価領域が全領域の25%程度に減少することもアイトラッカー



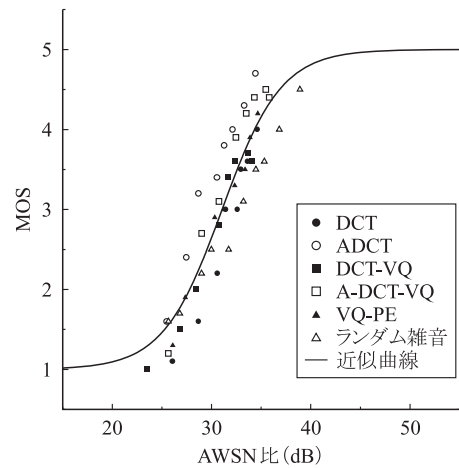
(a)テクスチャの複雑さを考慮しない場合



(a)テクスチャの複雑さを考慮しない場合



(b)テクスチャの複雑さを考慮した場合



(b)テクスチャの複雑さを考慮した場合

図8 MOSとAWSN比の関係(画像依存性) 画像の局所的な複雑さや平均輝度, 評価領域の推定を行うことで異なる画像でもMOSを推定しやすくなった。

図9 MOSとAWSN比の関係(雑音種類依存性) 画像の局所的な複雑さや平均輝度, 評価領域の推定を行うことで様々な劣化や雑音でもMOSを推定しやすくなった。

を用いた実験で明らかになったのでこのことも考慮に入れて, 最終的な評価システムを構築した。

その結果, 図8のように原画像が異なってもAWSN比とMOSの相関が0.78から0.93に高まるとともに, 異なる符号化雑音やランダム雑音に対しても図9のように0.82から0.91に相関が高くなることが明らかになった。

これらの研究成果を積み上げていくには3年という長い歳月を要したが, 一生懸命に研究に取り組む学生と350枚もの画像を主観評価してくれる学生が団結した成果と言える。

3. ‘ひと’の視覚特性から派生した研究

前章で述べたMOSを高い精度で推定できる客観評価値を求めるための研究は, 自身の生涯にわたる研究として現在も続けている^{(15), (16)}。しかし, 相関を0.7から

0.8にするのは比較的簡単でも, 0.8を0.9にすることは更なる積み上げが必要であり, まして, 0.9から0.95にするのは簡単にはできないのではない。

そこで, 軸足は主観画質ではあるが視覚特性に関連した派生研究から, 三つ研究を御紹介する。

3.1 画像の大きさが小さくなるとどうなるかという研究⁽¹⁷⁾

主観評価を行っているとき, 画面高 H の整数の距離で観視するのであれば, 劣化や雑音の見え方は画面の大きさには依存しないのかという質問を学生から受けた。当時はまだスマートフォンなど出現していない時代であったが, 液晶の密度が上がり始めていて, 小画面のものであればかなり高解像度な液晶ディスプレイが手に入るようになった。

そこで, 実際に画質を画面高135mmの中画面(Middle)と67.5mmの小画面(Small)を液晶モニターで主

観評価したところ、図 10 のように小画面にすると MOS が 0.5 ほど改善されることが分かった。しかも、この現象は、ランダム雑音を加えたときでもブロック雑音を加えたときでも同じような現象となったのである。

この現象を理解するために、過去の文献を調べたところ、SSF (Subjective Spatial Frequency) という概念があることが分かり^{(18), (19)}、画面が小さくなるとこれらの実験結果を説明できることが分かった。

簡単に言えば、画像を小画面上に表示すると、近くで見ることによる心理的な空間周波数特性が低域側にシフトし、中広域の雑音が見えにくくなることが証明できた。

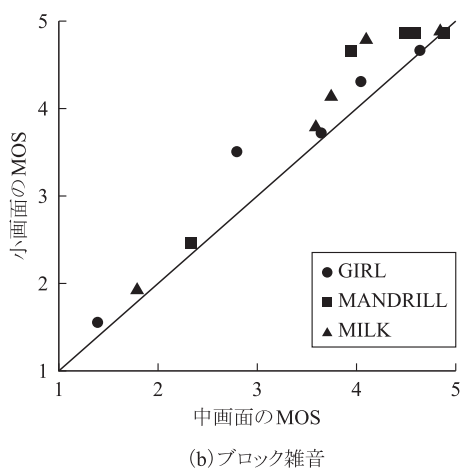
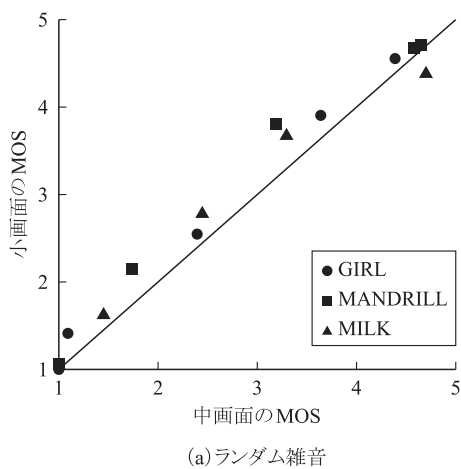


図 10 小画面と中画面の MOS の変化 ランダム雑音でもブロック雑音でも小画面の MOS が高くなるため測定点が実線の上に現れた。

3.2 動画に電子透かしを埋め込むという研究⁽²⁰⁾

動画符号化方式である MPEG の画質を考えていたときに、動きベクトルを故意に変更しても予測誤差が増えて符号量が増加するだけであることに気づき、透かし情報を P フレーム作成時に求まる順方向動きベクトルに埋め込む方法を考え、実際の動画像により性能を評価した。図 11 に実際の動画像シーケンス「Mobile」に 100 bit/frame の乱数データを埋め込んだ結果を示す。伝送レートが 2 Mbit/s 以上であれば、2% 以下の誤り率でデータを埋め込むことが分かった。

3.3 非線形処理によって超解像画像を作るという研究⁽²¹⁾

撮像系や表示系の解像度が 2K から 4K 更には 8K へと高くなるにつれて、それまで撮りためた画像を単純に拡大するのではなく、高解像度の画像を生成する超解像技術が必要となる。図 12 はそのための処理を表したもので、画像に含まれている空間領域の高周波成分を高域フィルタで抽出し、それを 3 乗演算して強調するとともに振幅制限を掛けた後に原画像に加えるという手法である⁽²²⁾。ここで重要なのは、「ひと」が注目する部分がどこなのかということと、その部分にどの程度の強調を掛けるのが最も画質が良くなったと感じるかどうかである。

そこで、画像を提示したときにどこに視点が集まるかという研究を始めた。図 13 は、アイマークレコーダ⁽²³⁾

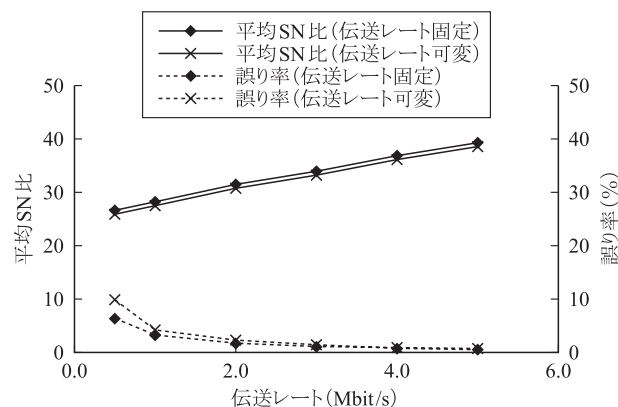


図 11 動画に 100 bit/frame の透かしデータを埋め込んだときの伝送レートに対する平均 SN 比と透かしデータ誤り率の変化 2 Mbit/s の伝送レートで、2% の誤り率で埋め込めることが分かった。

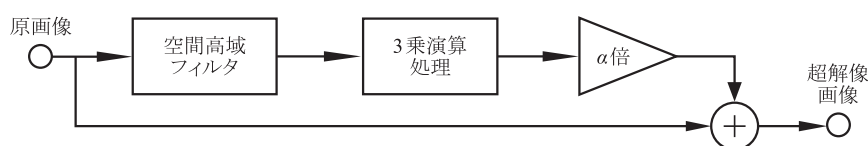


図 12 高空間周波数成分の強調による超解像処理法 高調波成分を加えることで超解像画像を得られるがどの領域にどの程度加えるかが画質を決める。



図 13 実験により求めた 4K 標準画像 Couple の注視領域 (視距離: 2H) 画像に応じた注視領域抽出と強調方法がこれからの研究テーマと考えている。

という装置を使って、4K 標準画像 Couple への視点分布を調べたものである⁽²¹⁾。この注視領域をどうやって予測するかが現在の私たちの研究の焦点になっているが、画像によっては注視点に特化した高域強調をするよりは、画像全体に強調信号を加えた方が主観的には良いことが分かっており、‘ひと’の視覚特性に関する新たな研究の始まりと考えている。

4. おわりに

JPEG や MPEG といった画像の符号化に興味を持ち、主観評価と客観評価の橋渡しをする研究を始めた 30 年前は全てが手探り状態であった。しかし、工学の基本である定量化の重要性を認識し、その困難な扉を少しずつ開けていくときの楽しさこそが研究の醍醐味といえる。また、一つのテーマにこだわっていると、たまには派生したいという欲求が生まれ、それが思わぬ研究結果につながることもあった。

未来を担う研究者の皆様の行く手は必ずしも楽しいことばかりではないかもしれないが、神様がほほえむときが必ず何回かはある。そのときに見過ごさないためにも不断の努力が必要だと思っている。

最後になりましたが、‘ひと’の主観画質に関する研究に携わり一緒に努力を重ねてくれた東京理科大学の学生諸君、有益な御助言を頂きました教員の皆様に感謝申し上げます。また、イメージ・メディア・クオリティ時限研究専門委員会発足に御尽力を頂いた皆様にこの場を借りて深謝致します。

文 献

- (1) A.N. Netravali and B.G. Haskel, Digital Pictures, Chapter 2, Plenum Press, 1988.
- (2) ISO/IEC FDIS15444-1, "JPEG 2000 image coding system," 2000.
- (3) ISO/IEC-13818, "Generic coding of moving pictures and associated audio information—Part 2: Video," 2013.

- (4) “デジタル映像の画質評価,” 山本英雄(編), 映情学誌, vol. 53, no. 9, pp. 1184~1208, 1999.
- (5) S. Hangai and K. Miyauchi, "Weighting function for evaluating coded picture quality in spatial frequency domain," Proc. PCS, vol. 90, no. 13, pp. 13.16-1-13.16-2, 1990.
- (6) S. Hangai, M. Ishikawa, and K. Miyauchi, "Coded picture quality evaluation in spatial frequency domain," Proc. PCS, vol. 3, no. 2, pp. 67-68, 1991.
- (7) S. Hangai, K. Suzuki, and K. Miyauchi, "Advanced WSNR for coded monochrome picture evaluation using fractal dimension," Proc. PCS, vol. 4, no. 2, pp. 92-95, 1994.
- (8) S. Hangai, T. Hamamoto, and H. Takano, "Collaborative method for quality evaluation coded pictures," Proc. IEEE ICIP, vol. 2, pp. 317-320, 1999.
- (9) A.N. Netravali and B.G. Haskel, Digital Pictures, Chapter 4, Plenum Press, 1988.
- (10) 半谷精一郎, 和田 徹, 宮内一洋, “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,” テレビジョン学会論文誌, vol. 46, no. 3, pp. 295-299, 1992.
- (11) 久保田啓一, 西澤台次, “テレビジョン系の 3 次元雑音評価関数とその高品位テレビへの応用,” 信学論(B), vol. J69-B, no. 5, pp. 503-511, May 1986.
- (12) 町澤朗彦, 吉田 實, “網膜モデルを用いた画質評価法,” 信学技報, IE86-92, pp. 9-16, 1986.
- (13) 草山貴由, 浜本隆之, 半谷精一郎, “人間の視覚特性を総合的に考慮した AWSNR の提案,” 映情学誌, vol. 55, no. 11, pp. 1443~1449, 2001.
- (14) 半谷精一郎, 岡本 淳, 宮内一洋, “画像の局所的なエントロピーを考慮した白黒静止画の画質評価尺度 WSNR の改良,” テレビ誌, vol. 49, no. 8, pp. 1078-1086, 1995.
- (15) 寺澤由明, 和田直哉, 半谷精一郎, “平坦領域の密集度を考慮したデジタル静止画像の客観評価尺度,” Proc. JQA2009, 2009.
- (16) 今井将太, 吉田孝博, 藤村嘉一, 半谷精一郎, “原画と劣化画を交互に表示したときの時空間周波数領域の雑音感度に関する検討,” Proc. PCSJ, 2012.
- (17) Y. Sugama, T. Yoshida, S. Hangai, C. Boon, and S. Kato, “A study on the subjective quality of pictures in small LCDs,” Proc. ICIP, pp. 131-134, 2005.
- (18) Y. Sugihara, A. Watanabe, and C. Yochida, “Subjective spatial frequency and picture quality estimating function I,” J. Inst. Telev. Eng. Jpn., vol. 41, no. 12, pp. 1173-1176, 1987.
- (19) Y. Sugihara, A. Watanabe, and C. Yochida, “Subjective spatial frequency and picture quality estimating function II,” J. Inst. Telev. Eng. Jpn., vol. 41, no. 12, pp. 1177-1182, 1987.
- (20) 菅原俊介, 和田直哉, 吉田孝博, 半谷精一郎, ブン チュンセン, “不適切な動きベクトルを用いる動画画像への情報理込方法とその抽出方法について,” 信学論 (D), vol. J91-D, no. 8, pp. 2014-2016, Aug. 2008.
- (21) 高橋駿輔, 増田恵子, 半谷精一郎, “注視傾向を用いた非線形超解像処理の検討,” Proc. PCSJ, 2016.
- (22) 合志清一, 寺川雅嗣, 三上 浩, 今井繁規, “非線形特性を応用した画像の超解像度化,” 情報科学技術フォーラム, vol. 8, no. 3, pp. 7-12, 2009.
- (23) ナックイメージテクノロジー, <http://www.eyemark.jp/>

(平成 29 年 1 月 25 日受付)



はんがい せいいちろう
半谷 精一郎 (正員: シニア会員)

昭 50 東京理科大・工・電気卒。昭 56 同大学院博士課程了。同年同大学助手、現在同大学教授。画質評価、話者識別、生体認証などの研究に従事。本会基礎・境界ソサイエティ英文論文誌編集幹事、同イメージ・メディア・クオリティ時限研究専門委員会委員長、通信ソサイエティユビキタスネットワーク社会におけるバイオメトリクスセキュリティ時限研究専門委員会委員長を歴任。著書に「コンピュータ概論」、「JPEG・MPEG 完全理解」など。