

動画像客観評価尺度 PQS_{video} -歪み要因の定義とその妥当性-

山下哲司 亀田昌志 宮原誠
北陸先端科学技術大学院大学、情報科学研究科
923-1292 石川県能美郡辰口町旭台 1-1
TEL 0761-51-1234 FAX 0761-51-1380
email tetsuji@jaist.ac.jp

Abstract

動画像の画質評価を、人の主観評価にかわって、客観的に評価できる尺度 PQS_{video} を作った。それは、MPEG などによって符号化、圧縮された画像に現れる歪みを客観的に評価しようとするものであり、原信号と復元信号の差信号で定義できる。ランダムノイズだけでなく、画質を大きく損なう local feature のすべてについて、視知覚的にどのような妨害を与えるかを考慮して、基本歪み要因を定義し、それを近似式で記述する。そして、これら基本歪み要因 (5つ) の主成分分析、重回帰分析による線形和によって得られる。本研究では各々の歪み要因の妥当性を検討するため実験を行い、歪み要因の定義の修正を行った。その結果、 PQS による主観評価の近似が重相関係数 0.996 まで向上し、主観評価を良く近似する評価尺度であることが示された。

An Objective Picture Quality Scale for Moving Image -Definition and Verification of Distortion Factors-

Tetsuji Yamashita, Masashi Kameda, Makoto Miyahara
Japan Advanced Institute of Science and Technology (JAIST)
School of Information Science
923-1292 Asahidai-1 Tatsunokuchi Ishikawa Japan
TEL 0761-51-1234 FAX 0761-51-1380
email tetsuji@jaist.ac.jp

Abstract

We have proposed PQS_{video} (Picture Quality Scale) to evaluate the coded moving image quality by an objective method. We have defined five distortion factors considering with human visual perception. The distortion factors must be defined considering not only global distortion but also the distortion for local features which is caused by each coding method. Applying both the principal component analysis method and the multiple regression analysis method to the obtained each distortion factor based on MOS (Mean Opinion Score), PQS_{video} is obtained. In this paper, we especially verified the definition of distortion factor of flicker and jadder. As the result, the weighted multiple correlation coefficient is improved from 0.975 to 0.996.

1 はじめに

ネットワークの普及やデジタル放送の開始にともない、画像の効率的な符号化の重要度は大きくなっている。このような中で、符号化、圧縮された画像の画質が受信者の要求に対して十分であるかどうかという問題が重要である。

従来、符号化画像の評価には主観評価が用いられることが多かったが、これは多くの労力と時間を消費する上、結果が不安定であるという欠点があった。また、従来用いられてきた客観評価尺度である平均二乗誤差やSN比などでは、符号化によって現れる local feature を損傷させる歪みを捕らえることが出来なかった。local feature を損なうような歪みは、ランダムノイズに比べ、非常に大きな妨害として知覚されるため、これらの方法で画質の正しい評価を行なうことは困難であった。

これを解決するために著者らは、主観評価を良く近似し、計算機で算出可能な客観的評価尺度 (PQS: Picture Quality Scale) を与えた。PQS は画像の符号化によって現れる一つ一つの local feature を損なう歪みが視知覚的にどのくらい妨害を与えるかを考慮して歪み要因を定義し、それらの線形結合で主観評価を近似するように作られた画質評価尺度である。

このような視知覚を考慮した歪み要因を研究していくことは同時に画像の特徴を損なう歪みとは何かを知ることと等価である。その結果、人が知覚する歪みを小さくする符号化アルゴリズムの開発やパラメータの決定を定量的、系統的に行うことができるようになる。静止画に対する客観的な評価尺度はすでに与えられている。[1][2]

動画像に関する客観評価尺度 PQS_{video} に対する提案は、これまでに行われてきており、フリッカ、ノイズ、ジャダー、動きぼけ、edge jitter に関して、基本歪み要因が定量的に定義された。しかし、これらの各基本歪み要因の定義についての詳細な検討はこれまで十分には行われていない。そこで本研究では、 PQS_{video} の基本歪み要因が妥当なものであるかどうかを実験を行うことによって検証し、問題点については再定義を行った。

2 PQS の基本構成

符号化によって生じる基本歪み要因 F_i を、原画像と符号化画像との差信号の関数で定義する。 F_i とし、符号化によって現れる local feature を損なう

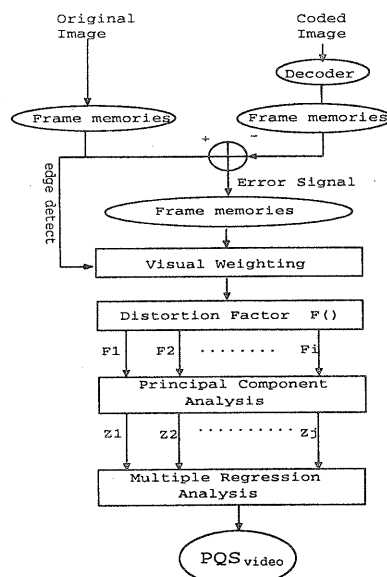


図 1: PQS_{video} の構成

歪みと、global な歪みがともに定義される。そして主成分分析によって F_i を無相関化した主成分を用いる。PQS は、この主成分の、主観評価 (MOS: Mean Opinion Score) に対する重回帰分析から、線形和の形で得られる。PQS 算出の説明図を図 1 に示す。

3 歪み要因とその定義

PQS_{video} における基本歪み要因 F_i は、原信号と符号化信号との差関数として定義される。しかし、目の複雑な動きを定式化することは非常に困難であるため、動きに対する視知覚の本質をとらえて動画画像特有の特徴から各 F_i を定義した。

動画像を見る場合、人の目が物体の動きに追従していれば、物体は網膜上に固定して捕らえられる。これは「追従視」と呼ばれ、妨害は静止画の場合と同様に感じられる。この点を考慮して、基本歪み要因が定義されてきた。

以下に先行研究 [3][4] によって定義された PQS_{video} における基本歪み要因とその定義式を示す。

3.1 歪み要因定義のための基本式

原画像を $X(l, m, n)$ 、符号化画像を $\hat{X}(l, m, n)$ とし、以下の式を定義する。

誤差画像：

$$e(l, m, n) = X(l, m, n) - \hat{X}(l, m, n) \quad (1)$$

weighted 誤差画像：

$$\bar{e}(l, m, n) = \mathcal{F}^{-1} \{ \mathcal{F} \{ e(l, m, n) \} R(u, v) \} \quad (2)$$

原画のフレーム差分画像：

$$d(l, m, n) = X(l, m, n) - X(l-1, m, n) \quad (3)$$

符号化画像のフレーム差分画像：

$$\hat{d}(l, m, n) = \hat{X}(l, m, n) - \hat{X}(l-1, m, n) \quad (4)$$

原画のフレーム差分のパワー：

$$D(l) = \sum_{m,n}^{all} \{ d(l, m, n) \}^2 \quad (5)$$

符号化画像のフレーム差分のパワー：

$$\hat{D}(l) = \sum_{m,n}^{all} \{ \hat{d}(l, m, n) \}^2 \quad (6)$$

(原画のフレーム差分のパワー)-(符号化画像のフレーム差分のパワー)：

$$D_e(l) = D(l) - \hat{D}(l) \quad (7)$$

式 (7) におけるフレーム差分 (フレーム差分の差分)：

$$\Delta D_e(l) = D_e(l) - D_e(l-1) \quad (8)$$

m : x 座標、 n : y 座標、 l :frame 番号、 $R(u, v)$:視知覚の空間周波数特性、 $\hat{\cdot}$:符号化画像、 $\bar{\cdot}$:weighted 画像

3.2 歪み要因 F_i

フリッカ フレーム毎に画質劣化の程度や性質が変化することがあり、これは輝度変化として知覚される。この、輝度変化はフリッカと呼ばれ、視覚的に大きな妨害となる。この視知覚特性は CFF(Critical Flicker Frequency) で表される。

ここでは、画面全体に現れる大面積フリッカについて定義した。フレーム差分の差分のデータ列をフーリエ変換し、CFF を乗じ、フリッカの検知限である 60Hz までの和をとった。

$$F_1 = \sum_{f=1}^{60} \{ |\mathcal{F} \{ \Delta D_e(l) \}| CFF(f) \} \quad (9)$$

ノイズ 符号化によるランダムノイズを評価する。ノイズは、静止画の場合のノイズの定義 (F_2) を用いる。[1] [2]

$$F_2 = \frac{\sum_{l=1}^A \sum_{m,n}^B \{ \bar{e}(l, m, n) T \}^2}{A} \quad (10)$$

$$T = \begin{cases} 1 & |\bar{e}(l, m, n)| > kth \\ 0 & |\bar{e}(l, m, n)| \leq kth \end{cases}$$

A は仮現運動の生じる範囲内、 B は全画面を示す。ジャダー ジャダーは同じフレームが繰り返されたときに起こる。動きに対しては、目は 5 Hz 程度までしか追従することができない。それ以上の動きには追従できず、動きの包括線のみを追従することになる。そこでフレーム繰り返しによってこの包括線からずれた変化がフリッカのように感じられる。このような妨害を「ジャダー」とよぶ。[5]

通常の動画像では動いている物体は、1 コマごとに大きく形を変えることはなく、1/30 秒単位で観察すればほとんど直線で一定の動きである。つまり、式 (8) は通常は 0 となる。一方ジャダーが発生して不自然な動きが起こればこの値は大きくなる。そこで、この式 (8) を用いてジャダーを定義する。

$$F_3 = \log \sum_{l=1}^{10} |\Delta D_e(l)| \quad (11)$$

動きぼけ 符号化された画像では、動いている像はぼけて見える。この動き像のぼけは、動きに目が追従している場合は、像が網膜上に固定して捕らえられるため、静止画がぼけた場合と同様に妨害を感じる。[6]

動きぼけを定義するために、動き領域についての原画像と符号化画像のスペクトル成分の差を求める。そして、この差分に視覚の空間周波数特性の weighting を掛けることによって動きぼけを定義した。追従視での視力は静止画に対する視力と変わらないので、weighting は静止画に対するものと同じものを用いることができる。

$$F_4 = \sum_{l=1}^C \sum_{u,v}^D \{ \{ |\mathcal{F} \{ X(l, m, n) M(l, m, n) \}| - |\mathcal{F} \{ \hat{X}(l, m, n) M(l, m, n) \}| \} R(u, v) \} \quad (12)$$

$$M(l, m, n) = \begin{cases} 1 & \text{動いている領域} \\ 0 & \text{動いていない領域} \end{cases}$$

Cは動きの生じている期間、Dは動きのシステムの周波数帯域の上限を示す。

Edge jitter ブロック処理を行なっている符号化画像ではエッジ付近での歪みが特に大きく、これがブロック状に広がっている。このエッジ付近での歪みがフレーム毎に変化してフリッカの様に知覚される。

Edge jitter は符号化画像のフレーム差分のエッジジッターの生じている領域 (Block 単位) に対してフーリエ変換をし、CFF を乗じ、その総和を取ったものとして定義した。

$F_5 =$

$$\sum_{block} \sum_{f=1}^{60} \left| \mathcal{F} \left[\sum_{m,n}^{block} \{d_e(l,m,n)E(l,m,n)\}^2 \right] \right| CFF(f) \quad (13)$$

$$E(l,m,n) = \begin{cases} 1 & \text{エッジジッターが検出された} \\ & \text{block の画素} \\ 0 & \text{エッジジッターが検出されない} \\ & \text{block の画素} \end{cases}$$

GはEdge jitterの生じているブロックすべてを表す。

4 歪み要因の検証

定義されたこれらの歪み要因が実際に人の視知覚特性を正しく反映したものであるかどうかを確かめるため、それぞれの歪みを故意に発生させたテスト画像を作成して、評価実験を行い、そこで得られたMOSと各 F_i の値との関係を観察することによって定義の妥当性を検討した。評価の方法は、表1に示すような二重刺激劣化尺度 (EBU法) による5段階劣化尺度法を用いた。

表 1: 評点と評価語

評点	評価語 (原画像と比較して)
5	妨害がわからない
4	妨害がわかるが気にならない
3	妨害が気になるが我慢できる
2	妨害が我慢できない
1	妨害が全く我慢できない

4.1 フリッカ

3章で定義された式(9)において、仮に、一定量のフリッカがフレーム毎に起こっているとすると、

輝度の増加と減少が打ち消しあい、 $\Delta D_e(l)$ は0になる。よって、この定義ではフリッカを正しく評価することができないと考えられる。そこでITU標準動画像シーケンスから、Flower GardenとBicycleを用いて評価実験を行った。原画像に対して、フレームごとに輝度を変化させてフリッカーを故意に発生させた。輝度の変化量および、変化パターンの異なる8種類の画像を用いて実験を行った。実験によって得られた F_1 とMOSとの関係を図2に示す。この図より F_1 とMOSはあまり相関がなく、式(9)による定義は適切でないと考えられる。

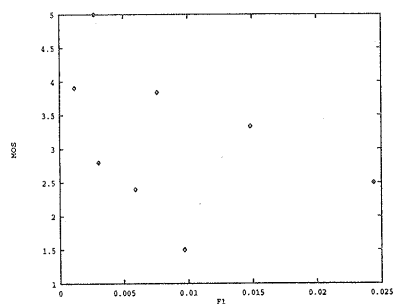


図 2: MOS と F_1 との関係 (改善前)

そこで $\Delta D_e(l)$ の代わりに $D_e(l)$ を用い、フリッカを再定義した。この式を以下に示す。

$$F_{1new} = \frac{\log \sum_{f=1}^{60} \{|\mathcal{F}\{HD_e(l)\}| CFF(f)\}}{I} \quad (14)$$

$$H = \begin{cases} 1 & \text{符号化画像の連続した2フレーム間のパワーの変化が原画よりも大} \\ -1 & \text{符号化画像の連続した2フレーム間のパワーの変化が原画よりも小} \end{cases}$$

Iは原画像の平均輝度である。

評価実験によって得られた F_{1new} とMOSの関係を図3に示す。この図より、MOSが小さくなる(妨害が大きくなる)と、評価値は大きくなっており、直線的に右下がりになっていることがわかる。よって定義された F_{1new} は主観評価を概ね良く表していると言える。

4.2 ジャダー

3章で定義されたジャダーの定義式(11)は「フレーム差分の差分($\Delta D_e(l)$)」の10フレーム分の和を計算し、それが最大となる所を、その動画像シーケンスのジャダー量としている。ところが、この方

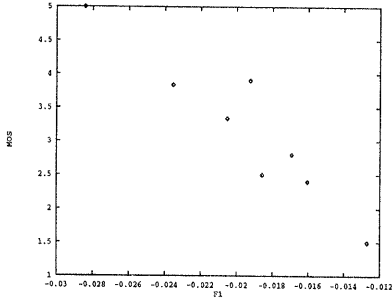


図 3: MOS と F_{1new} との関係 (改善後)

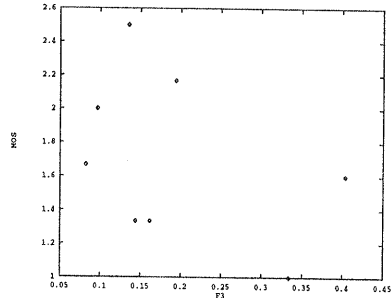


図 4: MOS と F_3 との関係 (改善前)

法では、視知覚の時間周波数的な特性が全く考慮されておらず、人の視知覚を反映した評価を行うことができないと考えられる。そこで、これを確かめるために評価実験を行った。テスト画像はITU標準動画画像シーケンスより、Mobile & Calendar と Bicycle を使い、この画像に対して、フレーム繰り返しを異なるパターンおよび回数で、故意に発生させたものを用いた。そして、得られたMOSとジャダー (F_3) との関係を観察した。結果を図4に示す。この図より、MOSと F_3 の間にはほとんど相関が見られない。そこでこれを改善するために再定義を行った。新しい定義では、視知覚の時間周波数特性を考慮するために、「フレーム差分の差分」のデータ列に対してフーリエ変換を行い、CFFを乗じ、和をとった。新たな定義式を以下に示す。

$$F_{3new} = \sum_{f=1}^{60} [F\{\Delta D_e(l)\} CFF(f)] \quad (15)$$

新たに定義された F_{3new} と MOS との関係を図5に示す。

この図より、劣化が大きくなるにつれて、 F_3 の値も大きくなり、その関係は線形に近いものとなっている。よって、 F_3 は視知覚を概ね良く表していると言える。

5 PQS_{video} と MOS の近似と考察

4章で示した再定義によって、PQS_{video} と MOS の近似がどのように変わったかを確かめるため、実際の符号化画像を用いて PQS_{video} 値を求め、MOS との比較を行った。

テスト画像には、ITU標準動画画像シーケンスより Flower Garden, Mobile & Calendar, Popple, Table Tennis, Bicycle を用いた、この画像データに M-C+DCT の符号化アルゴリズムを 4,9,15[Mbps] で適

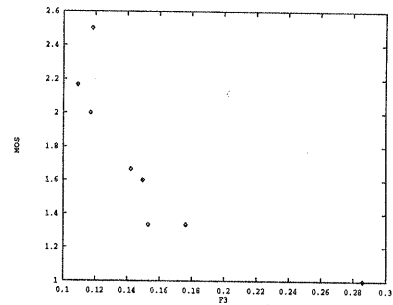


図 5: MOS と F_{3new} との関係 (改善後)

用した15種類の動画像と原画像との間の評価実験を行った。この結果得られたMOSとPQS_{video}の関係を図6に示す。PQS_{video}がどれだけMOSを近似しているかを定量的に表すため、重相関係数Rを次式によって求めた。

$$R = \frac{\sigma_{PQS}^2}{\sigma_{MOS}^2} \quad (16)$$

ただし、 $\sigma_{PQS}^2, \sigma_{MOS}^2$ は、

$$\sigma_{PQS}^2 = [E\{(PQS - \mu_{PQS})^2\}] \quad (17)$$

$$\mu_{PQS} = E\{PQS\} \quad (18)$$

$$\sigma_{MOS}^2 = [E\{(MOS - \mu_{MOS})^2\}] \quad (19)$$

$$\mu_{MOS} = E\{MOS\} \quad (20)$$

この結果、重相関係数が0.88となる。

このとき、図6においてMOSが2以下の、画質が悪い場合にPQS_{video}とMOSの近似が良くない様子が観察できる。これは、非常に画質が悪い場合にはジャダー (F_3) が大きく発生していて、他の

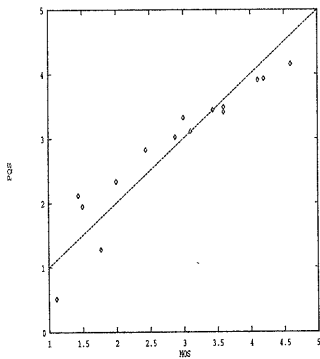


図 6: MOS と PQS_{video} (新) の関係

歪みがマスキングされるためではないかと考えられる。MOSにはマスキングにより、ジャダーによる歪みのみが反映されるが、 PQS_{video} はすべての歪みの線形和として表されるため、PQSがMOSよりも低いという現象が生じると考えられる。よって、画質が非常に悪い場合については別にする必要があると考えた。

そこで、 $MOS < 2$ のデータを除いて PQS_{video} を求めた。結果は図7に示すようになる。比較のため、旧定義の場合の結果も図8に示す。3.2で示した旧定義の場合は $R = 0.975$ 、新定義の場合は $R = 0.996$ となり、新しい定義でMOSが2以上の画像を用いた場合は大幅に近似が良くなった。

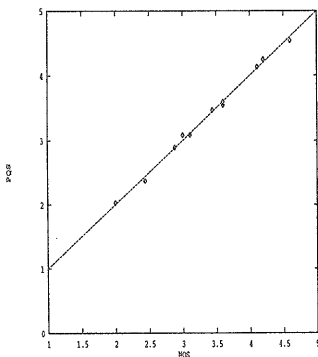


図 7: MOS と PQS_{video} (新) の関係 ($MOS \geq 2$)

6 まとめ

本研究では、人の主観評価を近似する、計算機で算出可能な動画客観評価尺度 (PQS_{video}) におい

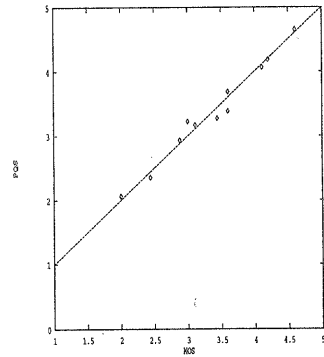


図 8: MOS と PQS_{video} (旧) の関係 ($MOS \geq 2$)

て、すでに定義された歪み要因について、実験によりその妥当性を検討し、視知覚特性を適切に反映したものであるかどうかを確かめた。そして、問題点については再定義を行い、より視知覚を反映するように改善を行った。さらに、新しく定義した歪み要因を結合することによって、PQS値を求めた。結果は、画質が非常に悪い場合を除いて、主観評価を良く近似しているといえる。今後、他の要因の定義の妥当性も検討するとともに、主観評価をさらによく近似する評価尺度を構築する予定である。また、非常に画質が悪い場合については、それが必要かどうかも含め、別途考える必要がある。

参考文献

- [1] Makoto Miyahara, Kazunori Kotani, V. Ralph Algazi"Objective Picture Quality Scale (PQS) for Image Coding"IEEE Trans. Commun., vol.46, NO.9 pp1215-1226 Sept. 1998.
- [2] 宮原、小谷、堀田、藤本."客観的画質評価尺度(PQS)-local featureの考慮と汎用性-",信学論 Vol.J73 B-1 No.3,pp.208-218,1990.
- [3] 宮原誠、堀田裕弘、豊芦穰、片山昌和、ムハマト ユスフエフェンディ,"動画の客観評価尺度-基本歪み要因の定義-",信学技報 CS95-147,IE95-99,pp1-6,1995.
- [4] 堀田裕弘、宮原誠、村井忠邦,"白黒動画における客観的画質評価尺度 (PQS_{video})",信学論, B-I,Vol.J80-B-I,No.6,pp.496-504,1997.
- [5] 宮原誠."動きの視知覚特性-フィールド繰り返しに起因する滑らかさの劣化",テレビ誌 Vol.27,No.1,pp24-30,1973.
- [6] 宮原誠."動画の画質-動き物体のぼけによるテレビジョン画質劣化-",テレビ誌 Vol.26,No.9,pp739-746,1972.