

## マルチメディアアプリケーション映像品質の自動客観評価法の精度改善の検討

Study on an improvement of estimation accuracy of objective video quality assessment for multimedia applications

杉本 修†  
Osamu Sugimoto酒澤 茂之†  
Shigeyuki Sakazawa小池 淳†  
Atsushi Koike

## 1. まえがき

携帯電話等の端末を対象としてマルチメディアコンテンツを提供する商用サービスが急速に普及している。これらのサービスの品質を監視する上で、利用者が主観的に感じる品質を自動的にかつ客観的に推定する必要が生じている。国際的な画質評価の専門家グループの VQEG では、マルチメディアアプリケーション映像品質評価方式の比較テスト(VQEG Multimedia Test [1])を実施しており、注目されている。

こうした背景から、筆者らは先に Full Reference 方式に基づくマルチメディアアプリケーション向けの映像品質評価方式を提案した[2]。先の提案では、圧縮符号化による劣化を前提とし、2種類のコーデックによりビットレート 128k~384kbps、で符号化した 3~15fps の CIF 画像の主観画質を相関 0.90 以上で推定可能であることを示した。本稿では、さらに QCIF 解像度の映像を評価の対象に加えるとともに、視覚特性を反映させたロジスティック関数によるスケールリングを映像特徴量の抽出の際に適用することにより、主観画質の推定精度を向上させる手法について検討する。

## 2. 従来方式の概要

従来方式では、圧縮符号化による劣化を前提とし、リファレンス画像(原画像)および評価画像(復号画像)を分析することにより3種類の映像特徴量を導出し、これらの重み付き和を求めた後、この値を評価映像のフレームレートに応じて補正することにより、客観評価値を導出している。なお、評価画像に対する主観評価値は、VQEG Multimedia Test のテストプラン[1]で規定されている ACR-HRR 法 (Absolute Category Rating with Hidden Reference Removal)に基づき測定される DMOS (Differential Mean Opinion Score)を用いる。

## 2.1. 従来方式による映像特徴量抽出法

提案方式による客観評価尺度  $Q_{obj}$  は、以下に示すパラメータ  $P_1, P_2, P_3$  の重み付け和により定義される。

$$Q_{obj} = \alpha P_1 + \beta P_2 + \gamma P_3 \quad (1)$$

ここで、最適な重みパラメータ  $\alpha, \beta, \gamma$  は、主観評価実験を用いたトレーニングにより導出するものとする。以下に従来方式による映像特徴量抽出法の概要を示す。なお、以下では、リファレンス画像を Ref 画像、符号化劣化を含む評価画像を Cod 画像と表記する。

## 1. ブロック歪度

フレーム内の隣接する画素ブロック間の DC 成分の差分をブロック歪みの知覚強度として定義する。ここでは、Ref/Cod 画像内の  $8 \times 8$  画素ブロックと図1に示す近傍 4 ブロック(当該ブロックの右隣、左下、下、右下)の DC 差分のフレーム内平均  $dDC_{Ref}(f)$ 、 $dDC_{Cod}(f)$  を Ref/Cod 画像から求め、さらに両者の差分のシーケンス内での最大値、最小値  $d_{max}, d_{min}$  を求める。最後に両者の差

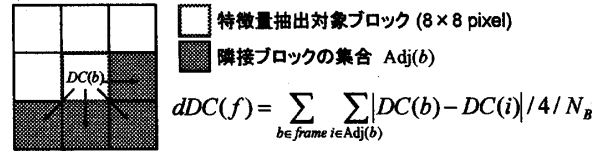


図1 隣接ブロックの定義

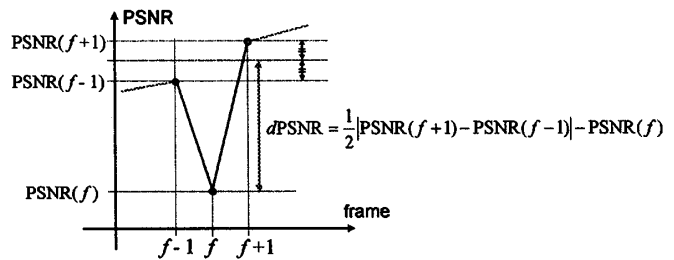


図2 時間的局所劣化度の定義

を式 4 に示すとおり対数化することにより、ブロック歪度  $P_1$  を得る。

$$d_{max} = \max\{f \in \text{sequence} | dDC_{Ref}(f) - dDC_{Cod}(f)\} \quad (2)$$

$$d_{min} = \min\{f \in \text{sequence} | dDC_{Ref}(f) - dDC_{Cod}(f)\} \quad (3)$$

$$P_1 = 20 \log_{10} \sqrt{\frac{255^2}{d_{max} - d_{min}}} \quad (4)$$

## 2. PSNR 値の時間的大域変動度

シーケンスの平均 PSNR は符号化品質の指標として一般的に用いられており、主観画質とも一定の相関があることが知られている。従来方式では、さらにシーケンス内での PSNR の時間的変動を反映するために、式 5 のとおり、各フレームの MSE のシーケンス内最大値、最小値のシーケンス内平均値からの偏りを求める。MSE の最大値、最小値、平均値をそれぞれ  $e_{max}, e_{min}, e_{ave}$  とすると、PSNR 値の大域的な時間変動度  $P_2$  は以下のとおり定義される。

$$P_2 = \log \left( \frac{e_{max} - e_{ave}}{e_{ave} - e_{min}} \right) \times e_{ave} \quad (5)$$

ここで、 $e_{ave}$  は、最後のシーケンス平均の MSE に応じて  $P_2$  の値を変化させるためのスケールリング項である。

## 3. PSNR 値の時間的局所劣化度

低ビットレート符号化では、キーフレームの挿入、シーンチェンジ、急激な動きの発生などによる PSNR の時間的な局所劣化が発生しやすい。そのため、この劣化を検出するための項を追加する。図 2 に示すとおり PSNR の V 字型の時間変化  $dPSNR(f)$  を検出し、そのシーケンス内最大値を時間的局所劣化度  $P_3$  として定義する。

$$P_3 = \max\{f \in \text{sequence} | dPSNR(f)\} \quad (6)$$

†株式会社 KDDI 研究所 映像通信グループ

2.2. 客観評価尺度の補正

前節に示した手順により客観尺度  $Q_{obj}$  と主観画質 DMOS の関係が得られるが、この段階では Cod 画像のフレームレートごとに若干異なる特性を示すことが知られている。具体的には、 $Q_{obj}$ -DMOS 特性を一次式で近似した場合、近似直線の傾きが異なるという傾向としてあらわれる。そこで、フレームレートに依存しない近似特性を求めめるために、図3のとおり、各フレームレートにおける近似特性を基準フレームレートの特性に変換する操作を行う。図3ではフレームレート b の映像の解析から得られた客観評価値  $Q_b$  をフレームレート a での特性において同一の DMOS 値を与える客観評価値  $Q_a$  に変換する例を示している。

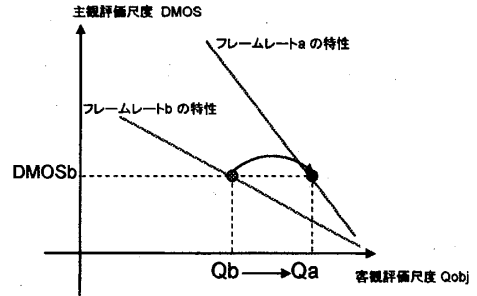


図3 フレームレートごとの  $Q_{obj}$ -DMOS 特性の補正

以上の補正により最終的に得られた客観評価値と DMOS の関係を3次の多項式により近似し最終的な  $Q_{obj}$ -DMOS 特性を求め、これにより主観評価尺度の推定値を求める。

3. 提案方式

3.1. 従来方式の課題

表1に示す条件で ACR-HRR 法に基づき主観評価実験を行って得られた DMOS 値と従来方式により求められた客観評価  $Q_{obj}$  の関係(CIF 解像度のみを対象)を図4に示す。同図において、絵柄ごとの推定精度を比較してみると、図中の丸で囲った絵柄 Susie の客観評価値と近似曲線の乖離が大きいがわかる。図5は、図4の丸で囲った条件のうち、上側のもの(絵柄 Susie に対して H.264 符号化、256kbps、15fps を適用した場合)の PSNR の推移を表したものである。このシーケンスの平均 PSNR は 38.3dB と比較的高いが、DMOS は 2.75 と DMOS 値の集合全体の中では大きく、主観評価としては低い結果となっている。これは、シーケンスの終盤、120フレーム付近の PSNR の下落が主観的印象に大きく影響を与えているためと考えられる。本来、このような PSNR の下落による主観画質の劣化は映像特徴量  $P_3$  で反映されるべきである。しかし、平均 PSNR が高いシーケンスでは、絶対的な PSNR の下落幅が小さくても、主観的には劣化が目立つため、 $P_3$  による劣化度は大きくなるべきである。しかし一方で、シーケンスの平均 PSNR が低い場合には、局所劣化度が前述の絵柄の PSNR の下落幅と同じであっても主観画質への影響はそれほど大きくないが予想される。従って、この現象を従来の映像特徴量  $P_3$  から検出するには限界があり、従来方式の主観画質推定精度を向上させるためには、評価画像の平均 PSNR の高低に応じて画像特徴量  $P_3$  の値を補正する必要があることがわかる。

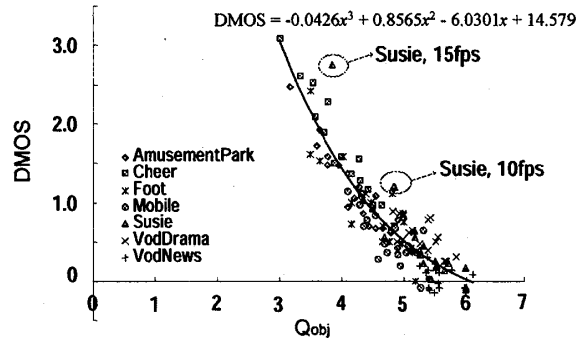


図4 従来方式による  $Q_{obj}$ -DMOS 特性(相関 0.9176)

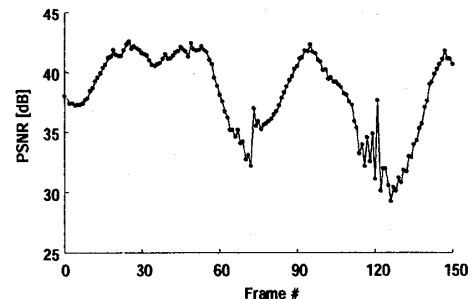


図5 Susie(H.264,256kbps,15fps)におけるPSNR推移

3.2. 提案方式による画像特徴量の補正

評価画像のシーケンスの平均 PSNR の高低に応じて画像特徴量の値を補正する方法としては、既に映像特徴量  $P_2$  に適用しているような平均 MSE または PSNR の値をそのまま乗じる手法が考えられる。この場合、平均 MSE/PSNR の値に応じて映像特徴量の値は線形に増加するような特性を持つことになる。一方、人間の視覚の特性を考慮すると、必ずしも MSE で示されるような誤差電力に線形に反応するとは限らず、ある範囲を超えた場合には評価値は飽和するような特性を持つことが考えられる。このような人間の知覚特性の近似にはロジスティック関数の適用が適している[3]ことが知られているが、提案方式では、ロジスティック関数の関数を1から引いた逆 S 字型の関数を映像特徴量  $P_3$  に乗じることで画像特徴量の値を補正する。

$$S_3(e_{ave}) = 1 - \frac{1}{1 + \exp\left(-\frac{e_{ave} - a}{b}\right)} \quad (7)$$

表1 実験条件

テスト画像	Amusement Park, Susie, Cheerleaders, Football, VoDDrama, VoDNews, Mobile and Calendar
映像フォーマット	CIF (352H×288V), 4:2:0, 15fps QCIF(174H×144V), 4:2:0, 15fps
シーケンス長	10sec
コーデック	H.264(QuickTime), Windows Media
Bitrate(bps), Framerate(fps) For CIF test	(128k,3), (128k,5), (256k,5), (256k,10), (256k, 15), (384k, 10), (384k, 15), (512k, 10), (512k, 15)
Bitrate(bps), Framerate(fps) for QCIF test	(32k,3), (32k,5), (64k,5), (64k,10), (128k, 5) (128k, 10), (128k, 15), (256k, 10), (256k, 15)
被験者数	36人
試験条件	ITU-T P.910 準拠

よって、映像特徴量  $P_3$  を導出する式 6 は以下のように変更される。

$$P_3' = \max\{f \in \text{sequence} \mid dPSNR(f)\} \times S_3(e_{ave}) \quad (8)$$

式7において、 $a, b$  は定数であり、最適化により相関が最も高くなる値を選択するものとする。図6は式7の特性を示したものである。式7は  $e_{ave} = a$  の点で 0.5 となり、その付近では線形の特性を示すが、 $e_{ave}$  の値の増加とともに 0 に漸近する。逆に、 $e_{ave}$  の減少時には、その値は 1 に漸近するという特性を持つ。すなわち、上述のとおり、 $e_{ave}$  の値がある範囲を超えた場合には、補正項が 1 または 0 に飽和するというように特性を反映することになる。

なお、映像特徴量  $P_2$  についても式 5 の右辺の第 1 項に式 7 と同様の特性の関数を乗ずることとする。ただし、 $P_2$  では  $e_{ave}$  の値の大小と飽和する方向が  $P_3$  とは逆になるため、

$$S_2(e_{ave}) = \frac{1}{1 + \exp\left(-\frac{e_{ave} - a}{b}\right)} \quad (9)$$

を乗じる。すなわち、提案方式の映像特徴量  $P_2$  の導出式は以下のとおりに変更される。

$$P_2' = \log\left(\frac{e_{max} - e_{ave}}{e_{ave} - e_{min}}\right) \times S_2(e_{ave}) \quad (10)$$

最終的に式 1 と同じように、映像特徴量  $P_1, P_2', P_3'$  の重み付き和を求め、これを客観評価指標  $Q_{obj}$  とする。

$$Q_{obj} = \alpha P_1 + \beta P_2' + \gamma P_3' \quad (11)$$

#### 4. シミュレーション実験

前述のロジスティック関数導入の効果を検証するために、シミュレーション実験を行った。主観評価データ取得の条件は、表 1 と同じとし、ロジスティック関数の適用の有無により主観画質推定の相関係数がどのように改善されるかを検証した。なお、ロジスティック関数の効果を比較するため、ロジスティック関数の代わりに  $S(e_{ave}) = e_{ave}$  という、 $e_{ave}$  に対して線形に増加するスケーリング項を適用した場合の相関係数も合わせて求めた。なお、式 11 の重み係数決定等の最適化は、CIF, QCIF 解像度ごとに個別におこなっている。

各条件下での相関係数の比較を表 2 に示す。映像特徴量  $P_2$  に対するロジスティック関数適用の効果を見ると、従来方式よりもロジスティック関数を適用した比較方式 1, 2 のほうが相関係数が向上していることがわかる。よって、特徴量  $P_2$  に対してはロジスティック関数適用の効果が確認できる。さらに、この条件下で特徴量  $P_3$  に対してロジスティック関数を適用した場合、CIF 解像度では 0.94、QCIF 解像度では 0.90 という相関係数を得る。よって、最終的に特徴量  $P_2', P_3'$  ともにロジスティック関数を適用することにより、主観画質との相関を最大にすることが可能であることが確認できる。このときの各解像度における  $Q_{obj}$ -DMOS 特性を図 7、図 8 に示す。図 7 より、CIF 解像度では当初の課題であった絵柄 Susie における近似曲線との乖離が解消されたことによる相関係数の向上が確認できる。一方、QCIF 解像度では、特に DMOS=0.5 以下となる評価画像において、 $Q_{obj}$ -DMOS 間の乖離が大きいため、CIF 解像度に比べると推定精度が下がっていることがわかる。これらの画像では、符号化劣化は発生しているものの、ポストフィルタ等の後処理による平滑化などで局所空間的な劣化は抑制される傾向にあるため、PSNR 値の低下の割には依然高い主観画質を得ている(すなわち、DMOS 値が小さくなる)ものと考えられる。また、こうした空間的に一様な劣化は二重刺激法では検出されやすいが、本実験では一重刺激法の一つである ACR 法に基づき主観評価尺度を求めているため、よりこうした傾向が現れるものと考えられる。

#### 5. むすび

マルチメディアアプリケーション映像品質の自動客観評価方式の精度向上のための改善手法について検討を行った。人間の視覚特性を反映させたロジスティック関数によるスケーリングを映像特徴量の抽出に組み込むことにより、CIF 解像度において相関係数 0.94 以上、QCIF 解像度については相関 0.90 以上で主観画質を推定することが可能であることを示した。

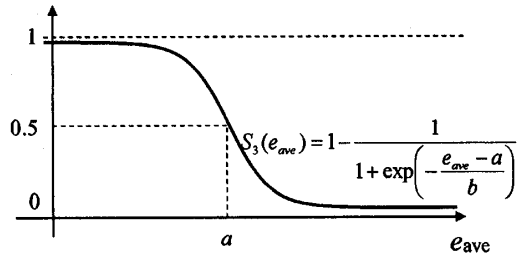


図 6 ロジスティック関数  $S_3(e_{ave})$  の特性

表 2 相関係数の比較

			相関係数	
	P2	P3	QCIF	CIF
従来方式	Linear	N/A	0.8727	0.9176
比較方式1	Sigmoid	N/A	0.8825	0.9198
比較方式2	Sigmoid	Linear	0.8866	0.9221
提案方式	Sigmoid	Sigmoid	0.9026	0.9425

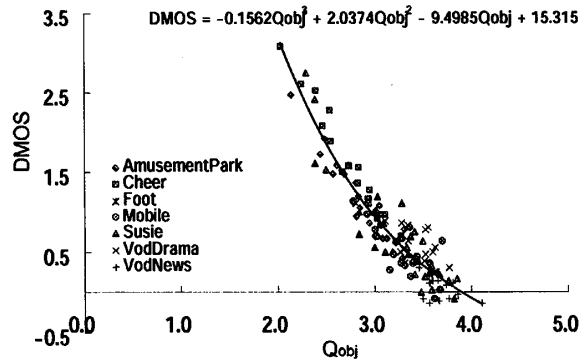


図 7  $Q_{obj}$ -DMOS 特性(提案方式、CIF 解像度;相関係数 0.9425)

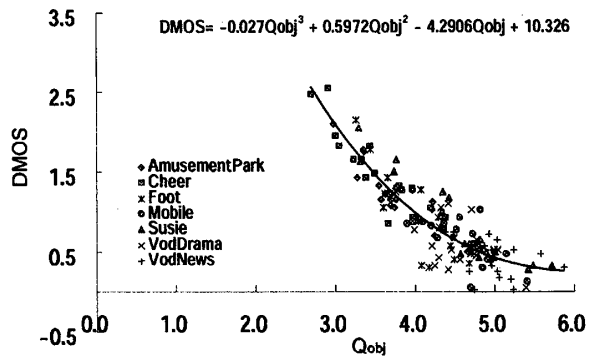


図 8  $Q_{obj}$ -DMOS 特性(提案方式、QCIF 解像度;相関係数 0.9026)

今後の課題としては、符号化劣化のみならず、ネットワークエラー等により映像劣化が発生した際の画質推定も可能なように提案方式を拡張することが挙げられる。

#### 参考文献

- [1] The Video Quality Experts Group, "Multimedia Group TEST PLAN (Version 1.11)", <http://www.its.bldrdoc.gov/vqeg/projects/multimedia/index.php>, Feb 2006
- [2] 杉本, 酒澤, 小池, "マルチメディアアプリケーション映像品質の自動客観評価法の一検討", 映像学 2006 年 年次大会, Aug 2006
- [3] 堀田, 宮原, 村井, "白黒静止画像における客観画質評価尺度の推定改善の検討", 信学論 B-I Vol. J80-B-I No.6 pp.505-514, Jun 1997