最適時空間解像度決定のための予測量子化パラメータに基づく主観画質推定

Subjective Video Quality Estimation based on Predicted Quantization Parameter to Determine Optimal Spatial and Temporal Resolutions

正会員 高木 基宏[†], 藤井 寛[†], 正会員 清水 淳[†]

Motohiro Takagi[†], Hiroshi Fujii[†] and Atsushi Shimizu[†]

Abstract Video quality improvement plays an important role in QoE in video applications. In this paper, we propose a subjective video quality estimation method that can compare subjective video quality in various spatial and temporal resolutions. Our method does not need to encode a video fully other than a few video encodings to estimate the relationship between bitrate and the quantization parameter.

キーワード: 主観画質推定, ACR 法, H.264/AVC

1. ま え が き

広域帯の無線回線や携帯端末の普及により、映像の視聴機会は大幅に増加し、不安定なネットワーク状況でも安定して、きれいな映像を配信する必要性が高まっている。近年では、ネットワーク状況に応じて映像を配信するために Http Adaptive Streaming (HAS) が普及してきている。HAS では、いくつかのビットレートの映像を用意するが、各ビットレートの解像度、フレームレートは固定である。しかし、映像ごとに最適な解像度、フレームレートは異なる。そこで本研究では、少ない計算量で最適な解像度、フレームレートを決定するための主観画質推定方法を提案する。

解像度、フレームレートの変化に対応した主観画質推定の多くは、Scalable Video Coding (SVC)¹⁾を対象としている。Sohn らは SVC で、解像度の変化の影響をエッジ特徴量で、フレームレートの変化の影響を動きベクトルで表現することで異なる解像度、フレームレートの映像に対して主観画質推定を可能にし、高い推定精度を実現している²⁾。しかし、主観画質推定に PSNR を用いており、復号映像が必要であるため、大量の映像を評価する際に符号化、復号のコストがかかる。Chien らは、符号化、復号の計算を避けるために、PSNR をビットレートとフレームレートから予測し、主観画質を推定したが³⁾、解像度の変化には対応していない。加えて、Sohn, Chien らの研究は解像度が QCIF

から CIF の範囲までとなっており、高解像度で適用可能かわからない。Li らは 4CIF, 60fps までの解像度、フレームレートで主観評価実験を行い、ビットレートから MOS を推定する rate-QoE モデルを提案している⁴⁾。高解像度、高フレームレートでの実験を実施しているが、SVC での配信を前提としており、任意の解像度、フレームレートに対応していない。

このように、従来研究では高解像度に対して主観画質の推定が可能で、かつ符号化、復号の計算量を削減した主観画質推定モデルとはなっていない。本研究では、符号化の計算量を抑えつつ、任意の解像度、フレームレートに対する主観画質推定方法を提案する。

2. 提 案 手 法

我々は文献 6) で量子化パラメータ (QP) は解像度、フレームレートごとに主観評価値と高い相関があることを利用して以下のように QP_q により主観評価値 s を近似した。

$$s = \alpha q + \beta \quad (1)$$

ここで α , β は解像度とフレームレートに依存する定数で、 q は各スライス・ヘッダに格納された QP の平均である。

文献 6) では複数の解像度、フレームレートで主観評価実験を実施し、得られた主観評価値から解像度、フレームレート別に回帰を行って α , β を求めた。しかし、この方法では主観評価実験を行った解像度、フレームレートでしか主観評価値を予測できない。そこで、本研究では α , β を解像度、フレームレートの関数として以下のようにあらわし、任意の解像度、フレームレートでの主観画質推定に対

2013 年 2 月 26 日受付, 2013 年 6 月 4 日再受付, 2013 年 7 月 3 日採録

[†]NTT メディアインテリジェンス研究所

(〒 239-0847 神奈川県横浜須賀光の丘 1-1, TEL 046-859-3173)

[†]NTT Media Intelligence Laboratories

(1-1, Hikarinooka, Yokosuka-shi, Kanagawa 239-0847, Japan)

応する.

$$s = \alpha(r, f)q + \beta(r, f) \quad (2)$$

$\alpha(r, f)$, $\beta(r, f)$ は以下であらわす.

$$\alpha(r, f) = a_1 \frac{r}{r_{max}} + b_1 \frac{f}{f_{max}} + c_1 \quad (3)$$

$$\beta(r, f) = a_2 \frac{r}{r_{max}} + b_2 \frac{f}{f_{max}} + c_2 \quad (4)$$

ここで, r は評価したい解像度での縦画素数×横画素数, f は評価したい映像でのフレームレート, r_{max} , f_{max} は主観評価実験を実施した解像度, フレームレートで最大のものである. a_1 , a_2 , b_1 , b_2 , c_1 , c_2 の係数は, 主観評価実験によって得た主観評価値と式 (2) による推定主観評価値の誤差が最小になるよう最適化を行うことで求められる定数である. これらの定数は映像に依らず同一である. したがって, 式 (3), 式 (4) の定数を一度決定すると, 式 (2) を用いて, 主観評価実験を実施していない任意の解像度, フレームレートに対して主観画質を推定できる.

式 (1), 式 (2) は QP から主観評価値を推定している. しかし, 固定ビットレートでの配信で本手法を適用するには, QP を求めるためにビットレート, 解像度, フレームレートの組合せの数だけ符号化することとなり, 計算量が膨大となる. そこで, 1 画素あたりの符号量から符号化前に QP を予測することで計算量を削減する. 本研究では, 1 画素あたりの平均符号量 b と QP q の関係は以下の式であらわす.

$$q = \epsilon \log(b) + \delta \quad (5)$$

ϵ と δ は映像に依存する定数である. QP が高くなると, ヘッダやモード情報などのオーバーヘッドの占める割合が多く, QP の変化に対する符号量の増減が小さくなることを式 (5) はあらわしている. 式 (5) は映像が同一であれば解像度, フレームレートが異なっても成り立つ. ϵ と δ は解像度, フレームレートの二つの異なる組合せで符号化して得られる b , q の組合せを用いて求める. 三つ以上の組合せで符号化し, 最小二乗法により求めても良い. b はビットレート, 解像度, フレームレートから符号化前に算出できる. 式 (5) により予測した QP を式 (1), 式 (2) に代入することで, 符号化前に主観画質を推定できる.

3. 評価と考察

表 1 の条件で主観評価実験を実施し, 得られた主観評価値を用いて評価を行った. 主観評価実験では, 7 種類の映像を 5 段階 ACR 法⁷⁾にて評価した. 一般家庭室内でのタブレット型端末を用いた映像視聴を想定し, 明るさは 200lux とした. また, 評価映像数が膨大なため, 被験者は評価を複数日にわたって実施した. 評価映像には ITE ハイビジョン・システム評価用標準動画第 1 版から Soccer, 第 2 版から Cosmos flowers, Woman at harbor, Studio concert, Horse racing を用いた. また, Astro Design の HDTV 評

表 1 主観評価実験条件
Experimental setting.

解像度	1920x1080, 1280x720, 960x540 640x360, 480x270
フレームレート (fps)	30, 15, 10
ビットレート (kbps)	15000, 8000, 4500, 2500, 1800 1200, 600, 400, 200
符号化方式	H.264 BaselineProfile, IPPPP
評価方法, 評価人数	5 段階 ACR 法, 32 名
ディスプレイ	9.7V 型液晶, 2048x1536x60Hz (iPad MC707 J/A)
視距離, 照度	4H, 200lux

表 2 映像別 MAD
MAD for each sequence.

映像	モデル 1	モデル 2
Soccer	0.12	0.23
Children playing	0.10	0.20
Torrent	0.72	0.38
Cosmos flowers	0.18	0.23
Woman at harbor	0.47	0.42
Studio concert	0.06	0.28
Horse racing	0.19	0.20
映像平均	0.26	0.28

価用動画像集 VT7001 から Children playing と Torrent のシーンを抜粋して用いた. これらの映像を表 1 の解像度, フレームレートにダウンサンプリングし, 各ビットレートで符号化することで評価用映像を用意した. インタレース/プログレッシブ変換には, Yet Another DeInterlace Filter (YADIF) を用い, フレームレートのダウンサンプリングは間引きによって実施した. 評価は, 主観評価実験に用いた映像の一つをテスト映像とし, 残りの映像をトレーニング映像とした. トレーニング映像の主観評価値を用いて式 (1), 式 (2) をそれぞれトレーニングし, テスト映像の主観評価値を推定した. これを各映像で繰り返し行い, 全映像の推定主観評価値を算出した. ここで, 式 (1) を特定の解像度, フレームレートごとに適用した主観画質推定をモデル 1 とし, 式 (2) を用いた任意の解像度, フレームレートに対応した主観画質推定をモデル 2 とする.

図 1 は各ビットレートで, 主観評価実験で得た主観評価値が最大となる解像度, フレームレートの主観評価値 (Optimal), モデル 1 の推定主観評価値が最大となる解像度, フレームレートの主観評価値 (Model1), モデル 2 の推定主観評価値が最大となる解像度, フレームレートの主観評価値 (Model2) の主観評価値である. Optimal には, 主観評価値の標準偏差をエラーバーとして示す. また, 映像別に Optimal の主観評価値と Model1, Model2 の主観評価値の差分の絶対値を各ビットレートで算出し, 平均した値 (MAD) を表 2 に示す. Torrent, Woman at harbor 以外はモデル 1 がモデル 2 と比べ, 最適な解像度, フレームレートでの主観評価値との差分が小さい.

ビットレート別に主観評価値の差をみてみると, モデル 2 では 600kbps と 4500kbps で最適解像度, フレームレートでの主観評価値との差が大きくなる傾向にあった. それ以

図 1 各ビットレートで最適な、または各モデルの推定主観評価値が最大となる解像度、フレームレートの主観評価値
MOS for the optimal/estimated resolutions and frame rates.

外のビットレートではモデル 1 よりも差分が小さい傾向にあった。これは、600kbps、4500kbps では推定主観評価値が実際の主観評価値よりそれぞれ高い値、低い値に推定されていたためである。高ビットレート、低ビットレートでは主観評価値が飽和し、QP の変化が主観評価値に与える影響も小さくなるが、その間のビットレートでは QP の変化が主観評価値に与える影響は大きく、かつ一定であるためと考えられる。提案法では、QP と主観評価値の関係を線形と仮定しているため、その変化を捉えられていないと考えられる。実際に、Woman at harbor では、600kbps で最適解像度、フレームレートは 640x360、10fps であるが、モデル 1 では 960x540、15fps、モデル 2 では 640x360、30fps と高い解像度、フレームレートを推定している。しかし、モデル 2 は任意の解像度、フレームレートに適用可能なことに加え、600kbps、4500kbps を除いた各ビットレートでの差分がモデル 1 より小さい傾向にあることから、解像度、フレームレートの決定に対して有用であるといえる。

4. む す び

ビットレートと QP、および QP と主観画質の関係を利用した主観画質推定方法を提案した。従来研究では、復号映像が必要、特定の解像度、フレームレートでのみ主観画質を推定可能と制限があったが、提案法では、ビットレートと QP の関係式を求めた後は符号化を行わず、任意の解像度、フレームレートでの主観画質推定を可能とした。提案法を、例えば映像のシーンごとに適用することで、主観画質の向上が見込める。しかし、任意の解像度、フレームレートに対応することで、最適解像度、フレームレートでの主観評価値との MAD が特定の解像度、フレームレートのみに対応した主観画質推定の MAD より大きくなった。

今後はオリジナル映像からの映像特徴量を式 (2) に取り入れ、精度向上に取り組む。また、解像度、フレームレートが動的に変化することによる画質変動の影響も検討していく。

〔文 献〕

- 1) "Advanced Video Coding for Generic Audiovisual Services", ITU-T Rec. H.264 and ISO/IEC 14496-10(MPEG-4 AVC), (2007)
- 2) H. Sohn, H. Yoo, W. De Neve, C. S. Kim and Y. M. Ro : "Full-reference video quality metric for fully scalable and mobile svc content", IEEE Trans. Broadcast., **56**, 3, pp. 269-280(2010)
- 3) M.-C. Chien, R.-J. Wang, C.-H. Chiu and P.-C. Chang : "Quality driven frame rate optimization for rate constrained video encoding", IEEE Trans. Broadcast., **58**, 2, pp. 200-208(2012)
- 4) M. Li, Z. Chen and Y.-P. Tan : "QoE analysis for scalable video adaptation", IEEE Conf. VCIP, pp. 1-6 (2012)
- 5) 高木 基宏, 藤井 寛, 渡辺 敬志郎, 岡本 淳, 清水 淳 : "主観画質に対する時空間解像度の影響", 2012 年映情学年次大, 3-8, pp.16-17(2012)
- 6) 高木 基宏, 藤井 寛, 清水 淳 : "ビットレートと量子化パラメータの関係を用いた主観画質推定", 2012 年映情学冬季大, 4-5(2012)
- 7) "Subjective video quality assessment methods for multimedia applications", ITU-T Rec. P.910(2008)

たかぎ もとひろ
高木 基宏 2009 年、慶應義塾大学理工学部管理工学科卒業。2011 年、同大学院修士課程修了。同年、日本電信電話（株）入社。現在、NTT メディアインテリジェンス研究所にて映像符号化の研究に従事。正会員。

ふじい ひろし
藤井 寛 1989 年、京都大学工学部情報工学科卒業。1991 年、同大学院修士課程修了。同年、日本電信電話（株）入社。以来、映像情報サーバ、マルチメディア情報流通方式、著作権管理方式、映像符号化の研究に従事。博士（情報学）。

しみず あつし
清水 淳 1990 年、日本大学理工学部電子工学科卒業。1992 年、同大学院修士課程修了。同年、日本電信電話（株）入社。ヒューマンインタフェース研究所およびサイバースペース研究所において低レート映像符号化方式や MPEG-2 や H.264 などの標準符号化方式の研究・開発に従事。現在、NTT メディアインテリジェンス研究所映像メディア符号化技術グループリーダー。正会員。