

振幅領域と空間周波数領域パラメータを使った ビデオ品質客観評価法

Method of objective evaluation of digital videos using parameters in amplitude and spatial frequency domains¹

池田宏明[†], 吉田 健[‡], 鹿島智子[‡]

Hiroaki IKEDA[†], Takeshi YOSHIDA[‡] and Tomoko KASHIMA[‡]

[†]千葉大学工学部

[†]Faculty of Engineering, Chiba University

[‡]千葉大学大学院自然科学研究科

[‡]Graduate School of Natural Sciences, Chiba University

要旨 – ビデオ送受信系末端間での客観的なビデオ品質評価法の開発を客観評価モデルは、過去の知見すなわちオリジナルビデオ(SRC)と処理済ビデオ(HRC)対とそれに対応した既知の主観評価値をモデルの教師データとする数学モデルで、モデルパラメータを、ある意味で最適に決定しようとする客観評価モデルについて述べている。具体的にこれを実行するために、既知の DMOS とともに SRC-HRC 対を非圧縮ビットマップフレームの時間系列として用意し利用した。なお、客観評価モデル開発後に評価対象とする未知ビデオ対およびその DMOS の推定に適用することを念頭に、可能な限り一般化した。

Abstract – This paper describes a model to estimate human subjective video quality scores based on a framework of double-ended with a full reference system or end-to-end network video systems where sending video and receiving videos are available. The model incorporates multi-dimensional and multi-variables all of which are objectively measurable. Its aim is to estimate subjective score in the difference mean opinion score between a reference video and corresponding degraded videos. The model is developed based on an application of the multivariable analysis and trained by a set of a priori known relationship between physically obtainable parameters and corresponding subjective score. The paper describes how the model incorporates a mathematical algorithm as an executable programme module, and in what degree the model can be applicable in order to account actual examples of video pairs for both training video set used and foreign video sets. Referring to an application of the developed model to a set of foreign videos, it reports the capability of the model to different quality range. In addition to the worked results, some issues for further study are highlighted.

Keywords: Objective video quality, network video, statistical estimation, multivariable analysis, DMOS, VQEG

1 はじめに

オリジナルビデオに関する完全な情報が、ある種の劣化を伴ったビデオとともに利用できるという枠組み、すなわち、ビデオ送受信系末端間での客観的なビデオ品質評価法の開発が、国連傘下の電気通信諮問委員会 ITU-R で 1997 年以来、継続的に研究されている。途中結果が 2000 年に公表され^[1]、また、ITU-R BT.2020-1 として刊行されている^[2]。そこでは、古典的な最大信号対雑音比 (PSNR) が有効な客観品質評価指標であるとされている。この研究課題は ITU-R SG 6 の Joint Working Party (JWP 6Q) に引き継がれ

た。これに呼応して、関連の専門家から成る任意団体 Video Quality Experts Group (VQEG) は、ビデオ品質の客観品質評価法の提案を、2002 年 11 月末を期限として広く募った。これを客観評価モデルと称する。客観評価モデルに期待されるのは、ビデオ主観品質と強く関連を持つ、あるいは主観品質を説明可能な実行可能プログラムである。この枠組みでは、所定のオリジナルビデオおよびそれを処理した(従って、ある種の劣化を伴う)ビデオの対について、モデル提案者とは独立に VQEG 内の独立研究機関で用意され、しかるべき主観評価が行われた。これらビデオ対とそれに対応した主観評価値は未公表とされ、全ての客観評価モデルが提出された後

¹ 同様の内容を Special session on visual communications and image processing conference (VCIP 2003), SPIE で発表予定。

にそれぞれのモデルの能力判定基準として使われた。このプロジェクトを VQEG FR-TV テスト第 2 フェーズと称する。著者らは、このプロジェクトに客観評価モデルの提案者として参加した。本稿では、そのモデル開発の考え方、技術的な内容およびその主観評価値推定能力について述べる。他機関から提案された客観評価モデルの性能との比較については、VQEG FR-TV テスト第 2 フェーズの最終報告を参照されたい^[3]。

このプロジェクトで平行して複数の評定者を使って実施された主観評価は、オリジナルビデオと処理済ビデオの両方を使うことから、ITU-R BT.500-11^[4]で勧告されている二重刺激連続品質尺度法 (DSCQS) が、VQEG が指定したルールに基づいて実施された。得られた主観評価値 Differential Mean Opinion Score (DMOS) の解釈等については、別稿を参照されたい^[5]。

筆者らの客観評価モデルは、過去の知見すなわちオリジナルビデオ (SRC) と処理済ビデオ (HRC) 対とそれに対応した既知の主観評価値をモデルの教師データとする数学モデルで、モデルパラメータをある意味で最適に決定しようとするものである。具体的にこれを実現するために、既知の DMOS とともに SRC-HRC 対を非圧縮ビットマップフレームの時間系列として用意し利用した。なお、客観評価モデル開発後にそれを評価対象とする未知ビデオ対およびその DMOS の推定に適用することを念頭に、可能な限り一般化した。

このように本稿では、ビデオ送受信系末端間での客観的なビデオ品質評価 (FR-TV) という枠組みで目標を実現するために、異次元・多変数を統合する方法を述べ、いかに最適化を行ったか、また、その結果の教師データ説明能力と未知ビデオ品質の説明能力について述べ、残された課題にも触れる。

2 モデリング方法

2.1 デジタルビデオからの客観値抽出の考え方

実際のデジタルビデオの情報量は膨大であり、理論的には必要な数の独立と考えられ

る物理量を抽出することが可能であろう。それを使って、対応するデジタルビデオ品質の主観評価値を説明することが可能である。しかし、その場合は、対象としたデジタルビデオとその主観評価値のみに有効で、他のデジタルビデオには通用しないなど、一般性を失うことになる。主観評価の客観評価モデルを構築するにあたって一般性を確保するには、DMOS が既知の多数の訓練ビデオが必要となる。一方では、利用できるビデオ対が有限で、場合によっては、主観品質が非常に高品質であったり、あるいは極端に低品質であったりする。現実問題としては、これまでに集積されている (標準) ビデオ対と DMOS 値を使う以外になく、その結果、予期しないビデオの客観品質値で主観品質を推定せざるを得ないことも覚悟しなければならない。これは、ここで述べる方法の限界となる。

デジタルビデオはカラー画像 (フレーム) の時系列であることから、個々のカラー画像に対して、二種類の特徴抽出法が考えられる。一つは空間振幅領域での特徴抽出、もう一つは空間周波数領域での特徴抽出である。また、これらは時間の関数であることを考慮すれば、さらに追加の特徴抽出が可能である。

2.2 空間振幅領域における説明変数抽出

SRC-HRC ビデオ対の対応するフレーム内の対応する画素間の誤差の二乗を計算する。通常、人間評定者は、画像フレームの中心部に主として注目すること等を考慮して、各画像フレームを図 1 に示したように、中心部分を含む 9 個のサブフレームに分割した。画素間の差の二乗平均を各ブロック毎に求め、9 個の独立説明変数とした。これを、カラーフレームの各チャンネルに適用した。カラー画像は、YCC や RGB などの独立した 3 チャンネルで構成されているから、合計で 27 変数となる。図 1 に示したように、それぞれの変数は時間軸方向に平均した。

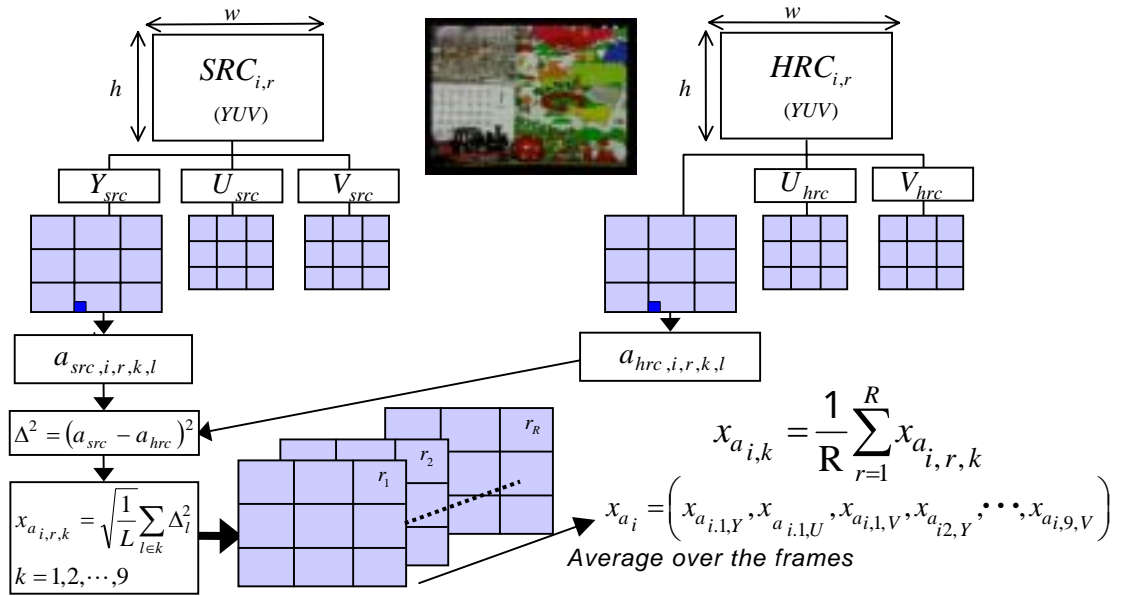


図 1 - 空間振幅領域における説明変数抽出

2.3 空間周波数領域における説明変数抽出

空間振幅領域における説明変数の抽出と平行して、各カラーフレームに 2 次元 wavelet 変換を適用して空間周波数領域における別な説明変数を求めた。本稿で述べる客観評価モデルでは、wavelet 変換はレベル 3 まで適用

し、図 2 に示したようにフレーム毎に 10 ブロックとした。これは、カラーフレームの各チャンネルに適用したので、合計 30 の独立と考えられる時間の関数である説明変数が得られる。これをここでは時間平均し、上で求めた空間振幅領域の説明変数と同時に用いることとした。

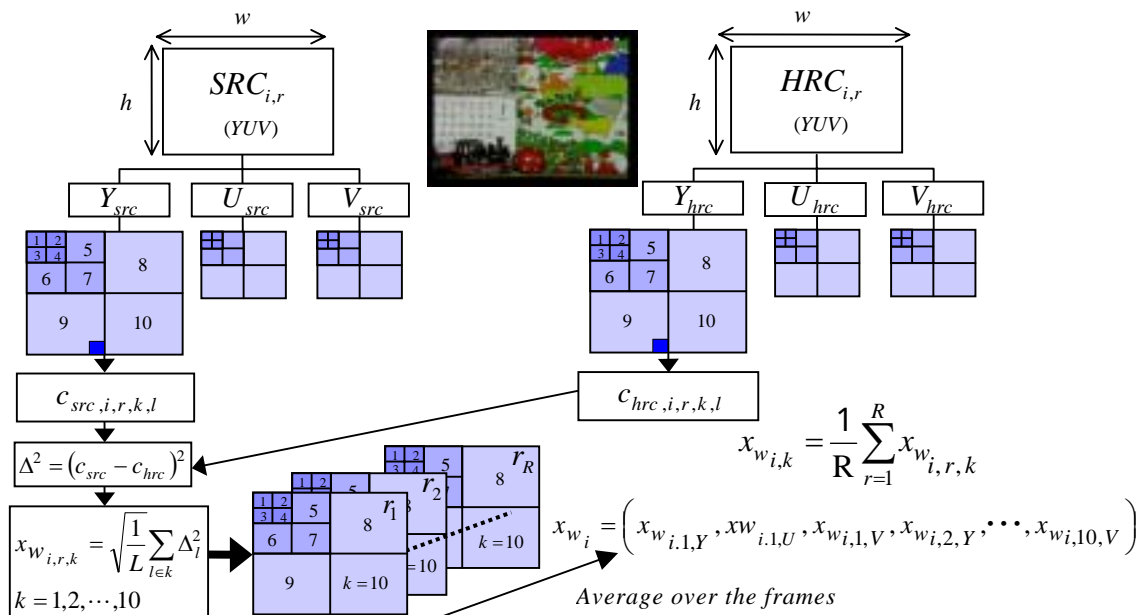


図 2 - 空間周波数領域における説明変数抽出

3 アルゴリズム

3.1 アルゴリズムの概要

入力される SRC-HRC ビデオ対から必要に応じて、必要な数の独立と考えられる説明変数を抽出可能であるが、どれほどの数のどれを説明変数とするかは、利用可能な DMOS 既知の訓練ビデオ対の DMOS の説明能力と同時に、DMOS 未知の新しい SRC-HRC ビデオ対の DMOS の推定が的確に行えるように決めなければならない。この最適化の問題は、必要な数の説明変数の選択と同時に解決しなければならない。

本稿では、SRC-HRC ビデオ対から得られる空間振幅領域における 9×3 個の説明変数と、空間周波数領域における 10×3 個の説明変数とから成る合計 $27 + 30$ 個の変数を、先ず最適化の初期変数集合とする。これを長さ $p = 57$ の説明変数ベクトルとする。このベクトルを DMOS 既知で利用可能な SRC-HRC ビデオ対から求めておき、客観評価モデルの訓練に使う。SRC-HRC ビデオ対の数を n とすれば、これらのベクトルが構成する行列は $n \times p$ の行列となる。

初期説明変数行列に多変量解析の手法である主成分分析を適用することにより、より少ない数の説明変数に対する重みの決定を以下に述べる手順で実施した。

3.2 主成分分析による説明変数の数の減少

上記 3.1 で導入した全ての説明変数ベクトルは互いに独立であると仮定しているが、客観評価モデルが未知のビデオ対に対しても DMOS 推定能力を有するように、より少ない数の説明変数で、主観評価の結果である DMOS を説明できるようにする、つまり、訓練ビデオ対に含まれるであろう個別の特徴の説明にのみ寄与する変数を取り去り、重要な変数のみを残すことによって、未知のビデオ対に対する説明能力を高め一般化しようとした。このために、以下の処理を導入した。

n 種類の SRC-HRC ビデオ対から得られる説明変数ベクトルの第 i 番目の正規化した説明変数ベクトルを $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ とし、

それで構成される行列を $\mathbf{A}_{n \times p}$ とする。すなわち、

$$\mathbf{A}_{n \times p} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i, \dots, \mathbf{X}_n)^t \quad (1)$$

ここで、正規化の方法は、平均値を 0 に、分散を 1 にするものとする。

行列 $\mathbf{A}_{n \times p}$ の相関係数行列を $\mathbf{B}_{p \times p}$ とし、その固有値を $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_j \geq \dots \geq \lambda_q \geq \dots \geq \lambda_p \geq 0$ 、 j -th 番目の固有値に対応する固有ベクトルが式(2)で表現されるものとする。

$$\mathbf{L}_j = (l_{j1}, l_{j2}, \dots, l_{jp}) \quad (2)$$

式(2)の各要素を重みとする重付き和である $z_{ij} = l_{j1}x_{i1} + l_{j2}x_{i2} + \dots + l_{jp}x_{ip}$ を要素とするベクトル、つまり元の説明変数の線形変換を要素とする新しいベクトルを $\mathbf{Z}_i = (z_{i1}, z_{i2}, \dots, z_{ip})$ として計算する。ここで、主成分分析の知見によれば、 $\text{Var}(\mathbf{Z}_i) = \lambda_i$ が成り立ち、 $\text{Var}(\mathbf{Z}_1) \geq \text{Var}(\mathbf{Z}_2) \geq \dots \geq \text{Var}(\mathbf{Z}_p)$ である。

この新しいベクトルの任意の組み合わせの相関係数はゼロで、これが n 種類の SRC-HRC ビデオ対の特徴を含んでいる。しかし、本稿では、さらに別な判定基準、すなわち、式(3)で定義した第 m 番目のベクトルの「寄与率」 c_m を導入して、さらに新しい説明変数ベクトルの数を減少しようと試みた。

$$c_m = \frac{1}{np} \sum_{i=1}^n z_{im}^2 \quad (3)$$

これは、次々に主要な固有ベクトルを採用することによって、どの程度の数(q)の新しい説明変数ベクトルが必要かを式(4)の条件で決めようとしている。

$$\sum_{m=1}^q c_m \geq 0.99 \quad (4)$$

値 q が決定できれば、 n 種類の SRC-HRC ビデオ対から得られた元の説明変数の線形変換として、式(5)が求められる。

$$\begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1q} \\ z_{21} & z_{22} & \cdots & z_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nq} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & \cdots & l_{q1} \\ l_{12} & l_{22} & \cdots & l_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1p} & l_{2p} & \cdots & l_{qp} \end{pmatrix} \quad (5)$$

式(5)を以下のように表現する。

$$\mathbf{Z}_{n \times q} = \mathbf{A}_{n \times p} (\mathbf{L}_1^t, \mathbf{L}_2^t, \dots, \mathbf{L}_q^t)$$

3.3 最小二乗法の適用

3.3.1 客観評価モデルの訓練

本稿で提案している客観評価モデルの訓練は、与えられた SRC-HRC ビデオ対に対応する既知の主観評価値 DMOS と行列 $\mathbf{Z}_{n \times q}$ のベクトル要素の重み付き和との差を最小二乗法により最小にするような重みを求めることである。ここで、単位縦ベクトルを行列 $\mathbf{Z}_{n \times q}$ に加えてオフセット項もあわせこむ。すなわち、式(6)の解を求めることである。

$$\|\mathbf{Z}_{n \times (q+1)} \mathbf{W}_{q+1} - \mathbf{D}_n\| \rightarrow \min \quad (6)$$

ここで、 n 次元ベクトル $\mathbf{D}_n = (d_1, d_2, \dots, d_n)^t$ は、式(7)で定義した主観評価値 mean opinion score (MOS) の差である DMOS を要素とする。

$$d_i = \text{MOS}_{\text{SRC}_i} - \text{MOS}_{\text{HRC}_i} \quad (7)$$

また、 $(q+1)$ 次元ベクトル

$\mathbf{W}_{q+1} = (w_0, w_1, w_2, \dots, w_q)^t$ は、式(6)を解くことによって得られる重みである。なお、要素 w_0 は客観評価モデルに導入されたオフセット値である。

3.3.2 未知の新しいビデオ対の DMOS 推定

新しい SRC-HRC ビデオ対に対する未知の主観評価値 DMOS を推定するためには、2.2 で述べた空間振幅領域における客観評価パラメータと 2.3 で述べた空間周波数領域におけるパラメータとを、全く同一の方法で抽出する。それを式(1)に示した行列に第 $(n+1)$ 行を加えた拡張行列 $\mathbf{A}_{(n+1) \times p}$ を得るために使う。これに続いて、3.2 と 3.3.1 に述べた方法を適用して、式(5)に対応する新しい行

列 $\mathbf{Z}_{(n+1) \times (q'+1)}$ を求める。ここで、 q' は必ずしも q と同じ値になるとは限らない。

更新された方程式(6)の解として求められる重みベクトル $\mathbf{W}_{q'+1}$ を用いて、式(8)のようにして DMOS の推定値 (VQR) が計算できる。ここで、 $\mathbf{Z}_{n+1} = (z_{n+1,1}, z_{n+1,2}, \dots, z_{n+1,q'+1})$ は DMOS 未知の、いま解析している SRC-HRC ビデオ対から抽出・変換された説明変数である。

$$\text{VQR} = \sum_{j=0}^{q'} z_{n+1,j} w_j \quad (8)$$

このように計算した VQR は、訓練ビデオ対に追加した DMOS 未知のビデオ対の主観評価値の推定値になる。残りの計算値 VQR は、客観評価モデルの訓練に使用したビデオ対の DMOS に対応する。その値は未知のビデオ対を加えたことによって多少値が変化するかもしれないが、大きな変化とはならないであろう。この観点から、提案したモデルは、過去に訓練を受けた内容に基づいて、それに対する影響を最小としながら新しい入力に対応するという意味で自立的である。

4 提案した客観評価モデルの実行結果

4.1 訓練ビデオ対に対する説明能力

提案した客観評価モデルの訓練の目的で、DMOS の値とともに公表されている VQEG FR-TV テストフェーズ 1 で使用された SRC-HRC ビデオ対を使用した^[1]。ビデオ対の総数は $(10 \times 16) \times 2 = 320$ である。

すでに報告してある通り^[6]、客観評価モデルの訓練だけを考えると、DMOS の説明能力すなわち DMOS と VQR の相関係数は、説明変数の数を増やすことによって任意に増大できる。その一例を散布図として図 3 に示した^[6]。一方では、訓練に使うビデオ対内の相関係数がある程度以上に増大させると、訓練に使っていない新しいビデオ対のあるべ

き主観評価の推定が大きく狂いその推定能力が減退する。これは、訓練ビデオ対に強く依存するためであり、いわば無意味なノイズを含めて訓練がなされたためである。

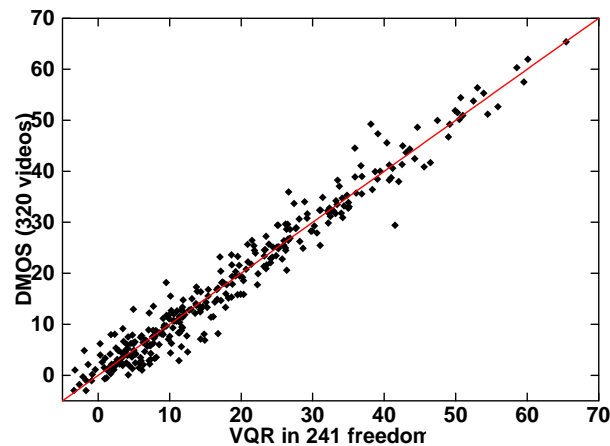


図 3 - 説明変数の数を多くした場合の散布図^[6]

このような望ましくない状況を避けるために、提案した客観評価モデルの訓練に関して、次のように考えた。訓練に使う DMOS の値は人間の主観であり、必ずしも決定論的な値ではなく、個人個人によって変わるいわば確率的な値で分散を有すること、従って、DMOS (平均値) だけではなくその分散も考慮して適度に訓練することとした。その結果の例を図 4 に示した。これは類似した未知のビデオ対の推定を可能とするいわば妥協であり、一般化になると考えた。

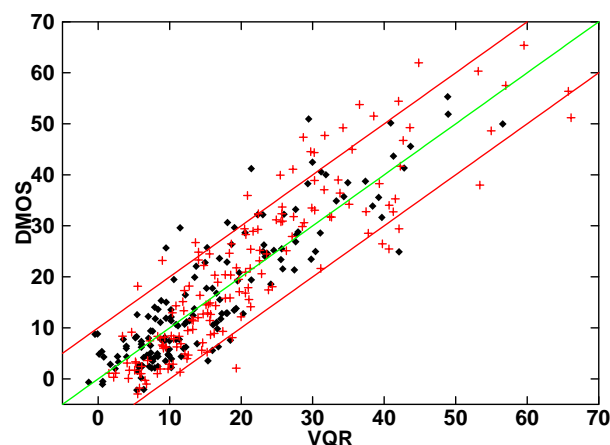


図 4 - 客観評価モデル訓練に用いた 320 ビデオ対の VQR と DMOS の関係 (黒は 525@60 Hz ビデオを赤は 625@50 Hz ビデオ)

4.2 未知ビデオ対の DMOS に対する説明能力 - 汎化能力

4.2.1 推定結果

VQEG FR-TV テスト第 2 フェーズでは、客観評価モデル開発を終了しそれを独立研究機関に提出後に、525@60 Hz ビデオ対と 625@50 Hz ビデオ対がそれぞれ 64 対の合計 128 ビデオ対としてモデル提案者に配布された。各モデルを実行した結果を二重に確認するためである。この時点では DMOS は未知のままである。その後、これらビデオ対に対応した DMOS がモデル提案者に通知された。この時点で、モデル提案者は各自のモデルの性能を確認できることになった。

図 5 と図 6 はそれぞれ、525@60 Hz ビデオ対と 625@50 Hz ビデオ対から推定した VQR と実際の DMOS との相関を示した散布図例である。いずれも説明能力が十分に発揮されていない。とくに大きな DMOS (より劣化の程度が大きい) においてこれが顕著である。

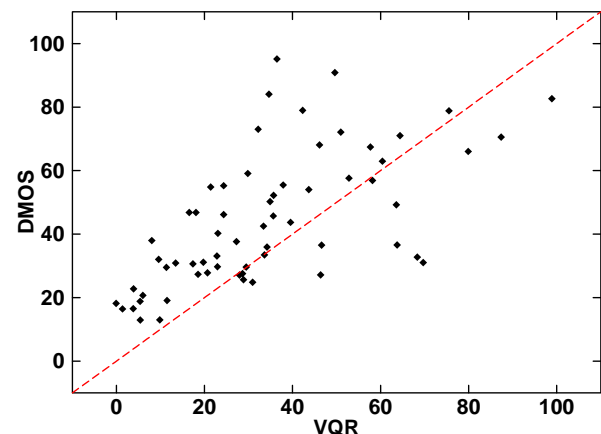


図 5 - 公表された DMOS とその推定値 VQR の関係(525@60 Hz ビデオ)

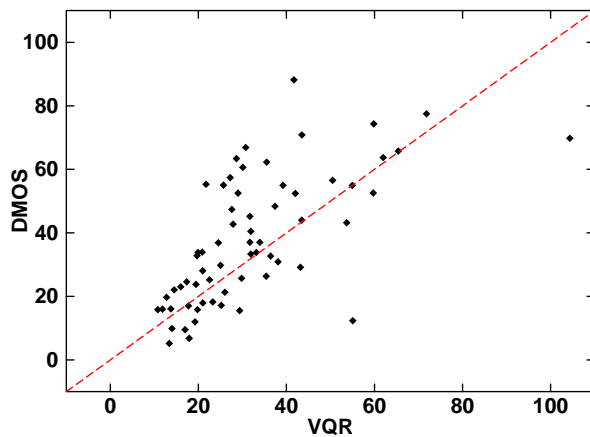


図 6 – 公表された DMOS とその推定値 VQR の関係(525@60 Hz ビデオ)

4.2.2 推定結果の考察

提案した客観評価モデルの訓練に使用した VQEG FR-TV テスト第 1 フェーズ^[1]のビデオ対と DMOS 推定対象とした VQEG FR-TV テスト第 2 フェーズ^[3]のビデオ対は, 基本的に独立であり, 主観品質評価値の範囲も同一ではない。第 2 フェーズでは, より劣化程度の大きなビデオも含まれていた。このため, いわば訓練済客観評価モデルによる外挿によって未知 DMOS を推定したことになる。これが説明能力の不足として図 5, 図 6 に現れたと考えられる。

VQEG FR-TV テスト第 2 フェーズでは, 複数の独立研究機関で実施された主観評価値の生データが最終的にモデル提案者に提供された。その解析・比較は別項^[5]に譲るが, 評定者個々の difference opinion scores (DOS) とその平均である DMOS の関係を, 図 7 と図 8 に示したように, かなり大きなばらつきのあるデータの平均値として DMOS が求められていること, 主観評価を実施した機関によって, 本来同一範囲の DMOS を期待したにもかかわらず偏りがあるなど, DMOS 推定の難しさが理解できる。すなわち, 図 7 では DMOS が [0, 80] であり, 図 8 では [10, 55] である。

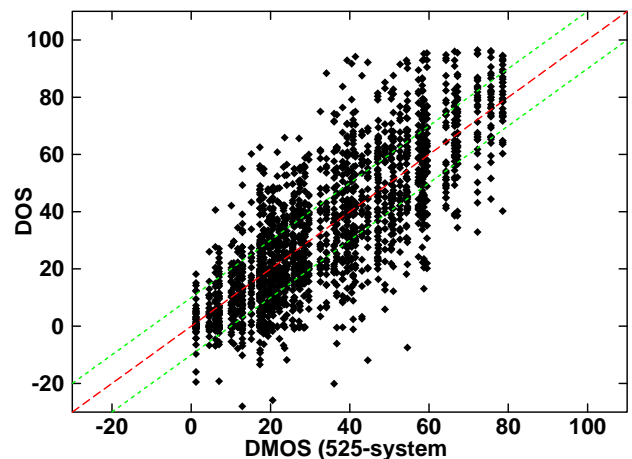


図 7 – DMOS の元となる DOS の分散 (525@60 Hz ビデオ)

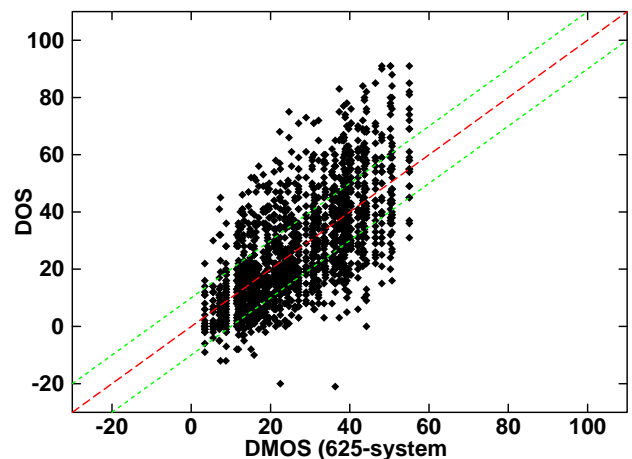


図 8 – DMOS の元となる DOS の分散 (625@50 Hz ビデオ)

VQEG FR-TV テスト第 2 フェーズ^[3]のビデオ対には, SRC-HRC の対応するフレーム間の時間的なミスマッチが存在していることが報告されている。提案した客観評価モデルでは, フレーム時間方向のレジストレーションは実施していないので, これも説明能力の欠如となった一因であろう。

5 おわりに

多変量解析の数学的手法を応用して, オリジナルビデオと処理済ビデオの主観品質の差である DMOS を客観的に推定するモデルについて述べた。この客観評価モデルを, DMOS が既知の両ビデオ対のデータから説明変数を抽出して訓練し, 未知のビデオ対にもこれを適用することにより, 自己増殖的に

DMOS に関する知識の吸収を可能とした。未知のビデオ対の主観評価値が訓練ビデオ対の主観品質の内挿である場合はこのモデルは良い推定を与え得るが，外挿となる場合は必ずしもよい推定を与えない。このための改良が今後の課題である。

今後のさらなる課題としては，1) SRC-HRC にフレームずれのあわせこみ前処理の導入，2) 第 2 フェーズの知見も加えて提案した客観評価モデルの再訓練をすると同時に，広い品質範囲のビデオにも適用可能とすることである。

謝辞：

三菱電機株式会社の的場氏にはモデル開発の初期の段階で援助を得た。また，VQEG FR-TV 試験フェーズ 1 で使用したビデオセットの使用許可をいただいた MR. PHILIP CORRIVEAU にも感謝する。

参考文献

1. Canada: Draft Video Quality Experts Group's Results, ITU-R JWP10-11Q/56-E (January 2001).
2. Report: Objective quality assessment technology in a digital environment, ITU-R BT.2020-1 (1999-2000).
3. ITU-R 6Q/146-E: Communication from VQEG: "Draft final report of FR-TV phase II validation test of objective measures of video quality for television using a full reference method"
4. ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures (June 2002).
5. 鹿島，青木，池田：ビデオ品質主観評価結果の比較検討，画像電子学会第 31 回年次大会 VMA 企画セッション(2003-06-20)
6. IEC TC100/PT62551: Preliminary study on objective measurement and assessment of video quality using wavelet transform, ITU-R SG 6 WP6Q/74-E (March 2002).
7. IEC 62551 Ed. 1.0: 2003, Multimedia systems – Quality assessment – Audio and video communication systems.