

フリッカーノイズの客観画質評価

武藤 圭吾[†] 甲藤 二郎[†]

[†] 早稲田大学理工学部コンピュータネットワーク工学科 〒169-8555 東京都新宿区大久保 3-4-1
E-mail: † {muto, katto}@katto.comm.waseda.ac.jp

あらまし 画質評価は、映像の劣化を測ることが目的である。画質評価には主観評価と客観評価があるが、客観画質評価は主観画質評価における煩雑な評価実験の手間を省くことができ、画像圧縮手法の研究にはかかせない尺度となっている。しかし客観評価尺度として広く用いられる PSNR(Peak Signal to Noise Ratio)は必ずしも主観評価結果との相関関係が高くないことは以前から指摘されている。その例として、DCT に起因するブロックひずみや、時間とともに画像の輝度値が変化することにより発生するフリッカーノイズがある。また近年研究が進められている客観評価尺度として SSIM(Structural SIMilarity)があるが、時間とともに発生するフリッカーノイズを反映させることはできない。本研究は特にフリッカーノイズについて評価を行い、既存手法である PSNR や SSIM にフリッカ評価量を組み入れることで主観評価結果との相関を高めることを目標とした。

キーワード 客観画質評価, PSNR, SSIM, フリッカー

The objective picture quality assessment of flicker noises.

Keigo MUTO[†] and Jiro KATTO[†]

[†] Department of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjyuku-ku, Tokyo, 169-8555 Japan
E-mail: † {muto, katto}@katto.comm.waseda.ac.jp,

Abstract Picture quality assessment tries to measure deterioration of the image. There are subjective assessment and objective assessment. The latter saves cumbersome tasks of the subjective assessment, and is very important for the studies of image compression. However, it has been pointed out that PSNR (Peak signal to Noise Ratio), which is widely used as an objective quality measure, has not sufficient correlation with the subjective quality assessment results. For example, there are blocking artifacts due to DCT and flicker noises caused by temporal changes of image luminance and color. SSIM (Structural SIMilarity) has been recently proposed as an objective quality measure, but it cannot reflect temporally oscillating flicker noises. This paper therefore focuses on flicker noises, and tries to improve correlations of PSNR and SSIM to subjective quality assessment results by incorporating flickering effect into the measures.

Keyword objective quality assessment, PSNR, SSIM, flicker noises

1. まえがき

動画はその必要な記憶容量の大きさから高能率圧縮を行うことが多い。その際効率的な符号化には劣化が伴う。画質評価はその劣化度合いを測るための非常に重要な技術である。画質評価は動画を視聴するユーザの主観評価が最も重要である。しかし効果的な主観評価には多くの評価者と評価環境が必要となる。そのためいつでも何度でも行えるわけではない。そこで動画の画素値や特徴量から画質評価を行う客観画質評価が重要となる[1]。

動画の客観画質評価としては下記式の PSNR(Peak signal to Noise Ratio)が使用されることが多い。また SSIM(Structural SIMilarity)と呼ばれる客観評価尺度が[2]により提案され、使用されることが増えてきた。

$$PSNR = 10 \log_{10} \left(\frac{255^2}{\frac{1}{N} \sum (F(x, y) - F'(x, y))^2} \right) \quad (1)$$

しかし動画において PSNR や SSIM は必ずしも主観評価とは一致しない。その原因のひとつに映像の明るさがフレーム単位で変化することによりちらついて見えるフリッカーノイズがある。PSNR は二乗誤差平均を規範とするため、量子化誤差の時間変動は反映しない。SSIM も同様にフレームごとに測定され、時間変動は反映しない。特にフレーム内符号化を動画に適用する場合や、複数の符号化を混在させる場合に強く現れる。筆者らによる複数の圧縮方式を用いた画像符号化手法[3]においても、フリッカーは前後フレーム

の合成により低減できることを確認しているが、主観による印象に比べPSNRには大きく影響しない。

そこで本稿ではフリッカーノイズについて、輝度値の時間的な変化から評価を行い、その評価量をPSNR、SSIM それぞれに組み入れたものとして新たな客観評価尺度を提案する。

2. フリッカー評価手法

2.1. フリッカーノイズ

フリッカーノイズは動画像の明るさが時間とともに増減することで、ちらついて見える現象である。符号化された動画像においてフリッカーノイズが現れる原因は、フレーム間で独立性が高い符号化が行われていることが原因である。例としてMotion JPEGやMotion JPEG2000のようなフレーム内符号化を動画像圧縮に適用する場合に顕著になる。中でも激しくない程度に動きがある領域において知覚されることが多い。これは静止領域ではフレーム間で独立した符号化を行っても変化が少なく、動きが激しい領域においてはフリッカーが発生しているとも知覚しにくく、主観評価には影響を与えないためである。一方でフレーム間差分を利用することでより効率的な符号化を行うMPEG、H.264/AVCなどではフリッカーは現れにくい。

フリッカーノイズは既存手法であるPSNR,SSIMでは測定が難しい。これはPSNRでは画素の差を二乗しているため値が増加しているか減少しているかについて注目が置かれていないことと、時間方向でどのような変化があったかを測定していないからである。SSIMも時間方向での変化については測定していないため、測定が難しい。

次に複数の動画像においてPSNR,SSIMと主観評価値の相関は次のようになる。動画像群はフリッカーのある動画も含まれる[3]の文献のものを使用した。

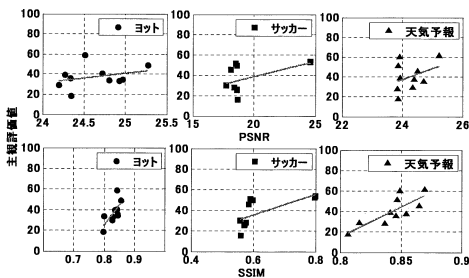


図1.PSNR,SSIMと主観評価値の分布,近似直線

主観評価値はITU-R BT.1788に記載されているSAMVIQ法によって測定されている。(PSNRがほぼ揃うように複数の圧縮方式をセットしたためではある

が)主観評価の幅の広さに対して、PSNRは非常に近い値が密集していることがわかる。これはフリッカーのようなノイズに対してPSNRが差をつけられていないことを示している。

2.2. 評価手法

本稿ではフリッカー評価に動画像の輝度成分を用いる。これは人間の視覚が輝度成分に敏感であるという点からである。

評価の手順は次のようになる。まず各フレームにおいて原画像 F と評価画像 G の輝度成分の差分を取り、差の小さいものの影響を下げるために二乗する、このときPSNRとは違い輝度差分の符号が失われないようにする。これは値の増減が明暗に関係してくるためである。これを輝度差分列 D とする。

$$D_n = \sum_{x,y} P(F_n(x,y) - G_n(x,y))^2$$

$$P = \begin{cases} 1 & F_n(x,y) - G_n(x,y) \geq 0 \\ -1 & F_n(x,y) - G_n(x,y) < 0 \end{cases} \quad (2)$$

フリッカーは前後のフレームに比べ輝度値が高い場合または低い場合に感知される。 n 番目のフレームにおいて、前後である $n-1$ と $n+1$ 番目のフレームの輝度差分値を用いてフリッカー評価値を定義する。

$$S_n = \left(D_n - \frac{(D_{n-1} + D_{n+1})}{2} \right)$$

$$Score = \frac{1}{\text{全画素数}} \sum S_n \quad (3)$$

実際に2つの動画像について評価を行った結果の D および S は図2,3のようになる。図2ではPSNRとは違い値の上下が輝度の上下に影響されるため、十分にフリッカーの有無が現れていると考えられる。

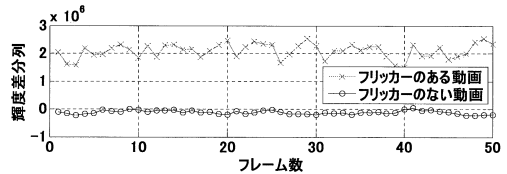


図2.フレームごとの D

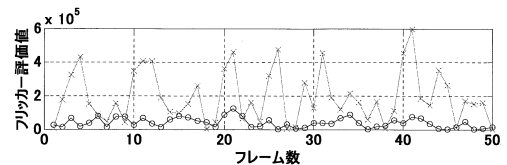


図3.フレームごとの S

次にフリッカー評価値 *Score* を用いて PSNR,SSIM に重み付けを行うことを検討する。重み付け方法については、フリッカー評価値に対して、係数 λ を用いて線形に重み付ける方式 ((4)式) と、PSNR と同じく対数をとって重み付ける方式 ((5)式) を検討した。

$$FPSNR = PSNR - \lambda \times Score \quad (4)$$

$$FPSNR' = PSNR - \lambda' \log(Score) \quad (5)$$

$$FSSIM = SSIM - \sigma \times Score \quad (6)$$

$$FSSIM' = SSIM - \sigma' \log(Score) \quad (7)$$

3. 評価実験

3.1. 評価対象

評価実験には前節でも使用した[3]の文献に用いられている動画像群を用いて行った。

まず動画像群について説明する。動画像は標準動画像 14 番ヨットハーバー(ゆっくりとしたパン)、20 番サッカー(パン)、39 番天気予報(背景 CG と人物の合成)を用いている。これら 3 種類に対して、Motion JPEG, Motion JPEG 2000, H.264(Intra Only)による符号化にて 3 種類、それらに動き補償の後、画素の重ね合わせによって量子化誤差の低減を図る方式(以下、画質改善手法)を加えたもの 3 種類、フレーム毎に複数の符号化方式を混在させ前記画質改善手法を加えたもの、および原画像で求めた動きベクトルを用いた画質改善手法の理想的な場合で、4 種類の計 10 種類の符号化方式を用いている。この動画像群に対する主観評価実験は SAMVIQ 法に基づき 15 人の被験者を対象に行っている。

この動画像群に対する PSNR,SSIM と主観評価の関係については図 1 で示している。主観的なフリッカーのおおよその目安は表 1 のようになる。H.264 はフレーム間差分を用いた動画圧縮特性が非常に優れているが、今回はフレーム内のみで行ったため、複数符号化混在の場合はフレーム間の独立性が高くフリッカーが強くなる。JPEG と JPEG2000 ではその性能の差から JPEG のほうが少し強くフリッカーが出ている。ここで[3]における画質改善手法とは前後フレームの情報を行くものであり、フレーム間の独立性が下がる傾向にある。そのため画質改善手法を加えた 4,5,6 はフリッカーが弱くなる傾向がある。また複数の符号化を混在させた方式とはフレーム毎に符号化方式を切り替えた動画像がベースとなる。そのためこちらも独立性が高い。改善手法には動きベクトル探索が用いられるが、圧縮方式を複数混在させた場合、動きベクトル探索に誤りが生じやすい、そこで動きベクトルに関してのみ理想的な値を用いたのが 9,10 となる。動画の

種類ではヨットハーバーがもっともフリッカーを感じやすく、ついで天気予報、サッカーとなる。これはサッカーでは動きが早くフリッカーを知覚しにくく、天気予報は画面半分ほどが静止領域であるためフリッカーが出にくいのである。

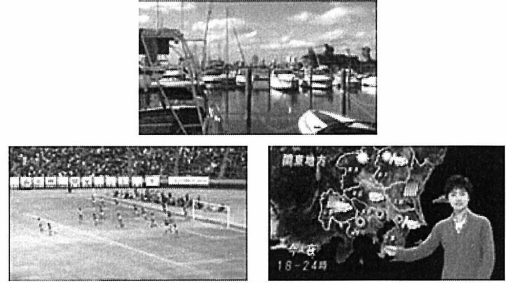


図 4.動画像サムネイル

表 1.主観的フリッカー強度

	ヨット	サッカー	天気予報
1.H.264/Intra	強い	弱い	強い
2.JPEG	中程度	弱い	弱い
3.JPEG2000	弱い	非常に弱い	非常に弱い
4.H.264/Intra 合成	中程度	弱い	中程度
5.JPEG 合成	弱い	非常に弱い	非常に弱い
6.JPEG2000 合成	弱い	非常に弱い	ほぼなし
7.二種混在合成	中程度	中程度	中程度
8.三種混在合成	中程度	弱い	中程度
9.二種混在理想	弱い	非常に弱い	弱い
10.三種混在理想	弱い	非常に弱い	非常に弱い

3.2. 実験結果

提案手法を用いてフリッカー評価を行った結果のフリッカー評価値を図 5 に示す。

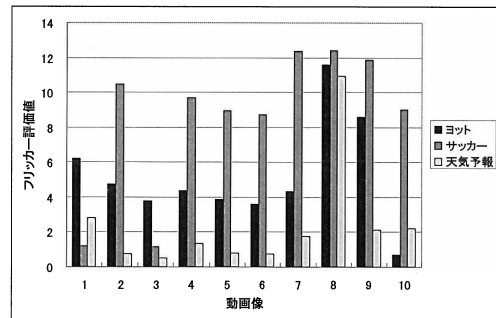


図 5.各動画像におけるフリッカー評価値

まず動画像の種類に応じたフリッカー評価値の順位について考える。ヨットハーバーと天気予報では、9種類においてヨットのほうがよりフリッカーが出ているという主観に沿った結果が出ている。サッカーについてはほかの二つと比べ高い値が多く出ている。しかし実際にはサッカーの動画はフリッカーを感じにくい。これはサッカーの動画像の輝度値の変化が少ないためではなく、動きによりフリッカーを知覚しにくいためである。そのため本節の提案手法だけでは動きのある動画像をうまく測定できず、動きによりフリッカー評価値に補正をかける必要があると考えられる。

次に各符号化手法におけるフリッカー評価値を考える。ここでは前述の理由より特にヨットと天気予報の二つに絞る。フリッカー評価値の低いものとして、2,3および5,6の天気予報の動画がある。これらは静止領域の影響からフリッカーが弱く、その点が評価値にも反映されたと考えられる。一方フリッカー評価値の多いものとして8,9がある。しかし視聴を行うと1の天気予報動画像や9の天気予報動画像でもフリッカーが強く発生していることから、10前後といった極端に高い値でなくとも、2程度ですでにフリッカーは出ていると評価されていることが考えられる。なお主観的にフリッカーがほぼ感じられない6の天気予報動画像のフリッカー評価値は0.73である。また天気予報動画は全体的にフリッカーが少ないと述べたが、1では高めに評価されている。これは1のH.264(intra)では画面内予測と呼ばれる、フレーム内で類似する領域から予測を行うため、静止領域においても動領域の影響を受けフリッカーが発生していると考えられる。4,5,6および9,10は1,2,3および7,8を改善した方式である。これによりフリッカー評価値も7,9およびもともと評価値の低いものを除き概ね減少している。

3.3. PSNRへの重み付け

フリッカー評価値を用いて式(4)(5)によりPSNRへ重み付けを行う。それぞれの式について λ と λ' の最適な値を求め、その値による重み付けの効果を確認する。最適値は λ を0~2の間を変化させ、動画像の種類ごとに重み付けされたFPSNRと主観評価値について最適な相関を出す値を調べる。ここで相関係数には一般相関係数に加え、値の順位のみで相関係数を求めるSROCC(Spearman's rank correlation coefficient: スピアマンの順位相関係数、以下SROCC)を用いる。

まず線形に重み付けを行った式(4)の結果を図6,7に示す。 $\lambda=0$ の時点は重み付けがないため、PSNR単体での相関となる。

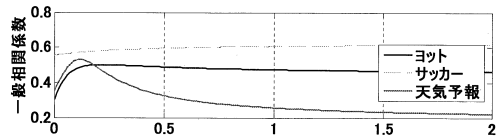


図6.線形重み付けによる一般相関係数の変化

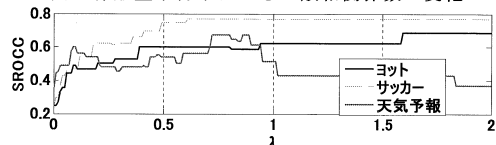


図7.線形重み付けによるSROCCの変化

一般相関係数はサッカーに関しては λ の増加に伴い増加し続け、フリッカー評価値単体との相関であるまでそれぞれ増加する。ヨットは $\lambda=0.22$ の時にピークを向かえ、相関係数は0.502となる。天気予報は $\lambda=0.12$ の時にピークを迎え、相関係数0.532となる。SROCCはサッカーとヨットに関しては λ の増加に伴い同様に増加する。天気予報は $\lambda=0.72$ の時にピークを迎えSROCC0.672となるが、 $\lambda=0.09$ においてもSROCC0.600と高い値を出す。

次に対数を用いて重み付けを行った式(5)の結果を図8,9に示す。

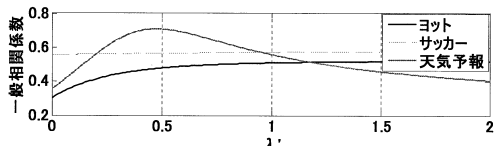


図8.対数重み付けによる一般相関係数の変化

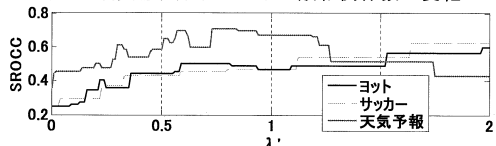


図9.対数重み付けによるSROCCの変化

一般相関係数は、ヨットとサッカーに関しては λ' の増加に伴い増加し続けフリッカー評価値単体との相関である相関係数まで増加する。天気予報は $\lambda'=0.48$ の時にピークを迎え、相関係数0.709となる。SROCCは、サッカーとヨットに関しては λ' の増加に伴い同様に増加し続ける。天気予報は $\lambda'=0.73$ の際にピークを迎えSROCC0.709となるが、 $\lambda'=0.60$ 前後においても0.697と高い値となる。

これらの結果から、各動画像および全体での最適な λ, λ' を以下のように定める。値の上昇が止まらない場

合には、1.00 とした。

表 2. 最適な λ, λ'

	ヨット	サッカー	天気予報	全体
λ	0.22	0.61	0.12	0.17
λ'	1.00	1.00	0.48	0.60

求められた λ, λ' と式(4)(5)を用いて相関係数をまとめると表 3 のようになる。FPSNR と FPSNR' による相関係数の上昇を見ると、必ずしも一方が有利というわけではないが、PSNR よりは性能が上昇していることがわかる。

表 3. FPSNR の相関係数

	ヨット		サッカー		天気予報	
	一般相関係数	SROCC	一般相関係数	SROCC	一般相関係数	SROCC
PSNR	0.305	0.248	0.557	0.248	0.357	0.358
個別 λ, λ'						
FPSNR	0.501	0.503	0.600	0.770	0.533	0.564
FPSNR'	0.510	0.467	0.571	0.467	0.709	0.588
全体 λ, λ'						
FPSNR	0.500	0.467	0.578	0.539	0.509	0.539
FPSNR'	0.490	0.503	0.567	0.455	0.689	0.697

各動画における全体での λ, λ' を用いた場合の FPSNR, FPSNR' と主観評価値の分布を図 10 に示す。図 1 で示した PSNR との分布と比較すると、近時直線に近づいている様子がわかる。

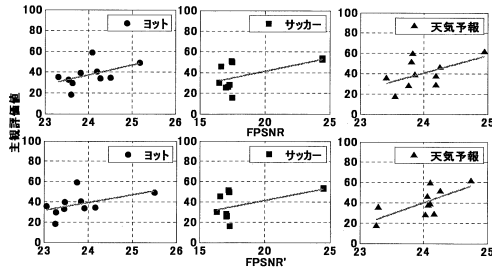


図 10. FPSNR, FPSNR' と主観評価値の分布と近似直線

3.4. SSIM への重み付け

次に SSIM に対して重み付けを行う。SSIM は PSNR に比べ値が小さいため、 σ を 0~0.2 の範囲に設定した。まず線形に重み付けを行った式(6)の結果を図 11,12 に示す。

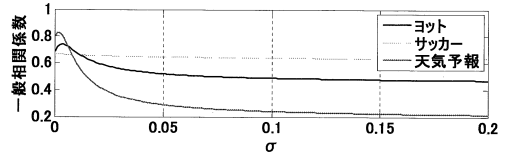


図 11. 線形重み付けによる一般相関係数の変化

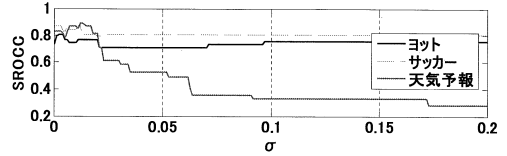


図 12. 線形重み付けによる SROCC の変化

一般相関係数はサッカーに関しては σ の上昇に伴い緩やかに減少した。ヨットは $\sigma=0.003$ の時にピークを迎え、相関係数 0.743 となる。天気予報は $\sigma=0.002$ のときにピークを迎え、相関係数 0.829 となる。SROCC においてもサッカーは同様に減少した。ヨットは $\sigma=0.002$ のときにピークを迎え、SROCC 0.806 となる。天気予報は 0.012 の時にピークを迎え、SROCC 0.890 となる。

次に対数を用いて重み付けを行った式(7)の結果を図 13,14 に示す。

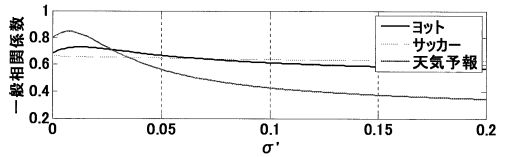


図 13. 対数重み付けによる一般相関係数の変化

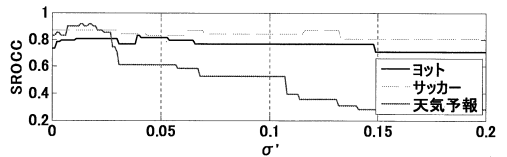


図 14. 対数重み付けによる SROCC の変化

一般相関係数はサッカーに関しては線形の時と同様に σ' の上昇に伴い緩やかに減少した。ヨットは $\sigma'=0.013$ のときにピークを迎え、相関係数 0.730 となる。天気予報は $\sigma'=0.007$ のときにピークを迎え、相関係数 0.847 となる。SROCC はサッカーについては減少と増加を繰り返したが、 $\sigma'=0$ の時以上の結果はでない。ヨットは $\sigma'=0.039$ の時にピークを迎え、SROCC 0.830 となる。天気予報は $\sigma'=0.013$ の時にピークを迎え、

SROCC 0.915 となる。

これらの結果から、各動画像および全体での最適な σ, σ' を以下のように定める。

表 4.最適な σ, σ'

	ヨット	サッカー	天気予報	全体
Σ	0.003	0	0.002	0.0025
σ'	0.013	0	0.007	0.010

求められた σ, σ' と式(6)(7)を用いて相関係数をまとめると表 5 のようになる。サッカーにおいては値の上昇は望めないが、他の二つについては SSIM よりも性能が上昇していることがわかる。

表 5.FSSIM の相関係数

	ヨット		サッカー		天気予報	
	一般相関係数	SROCC	一般相関係数	SROCC	一般相関係数	SROCC
SSIM	0.687	0.733	0.666	0.867	0.802	0.830
個別 σ, σ'						
FSSIM	0.742	0.806	0.666	0.867	0.829	0.830
FSSIM'	0.730	0.806	0.666	0.867	0.847	0.903
全体 σ, σ'						
FSSIM	0.741	0.806	0.665	0.867	0.824	0.830
FSSIM'	0.729	0.794	0.660	0.867	0.843	0.903

各動画像における全体での σ, σ' を用いた場合の FSSIM, FSSIM' と主観評価値の分布を図 15 に示す。PSNR に比べると図 1 からの変化は少ないが、天気予報では近似直線に近づいている様子がわかる。

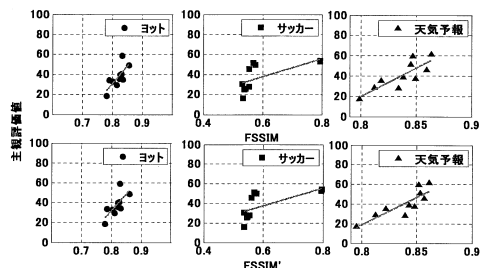


図 15. FSSIM, SSIM' と主観評価値の分布と近似直線

4. まとめ

本稿では客観画質評価尺度として、動画像のフリッカーノイズを評価し、PSNR と SSIM へ重み付けする手法を提案し、主観評価との相関が強くなることを確認

した。動画より最適な重み付け係数に差があるため、動画ごとでの最適な係数を見つけることが課題となる。その際、サッカー動画像のような動きの激しい動画像に対応するために動画像の動きを測定し反映させる必要があると考えられる。

文 献

- [1] Video Quality Experts Group (VQEG): <http://www.its.bldrdoc.gov/vqeg/>.
- [2] Z.Wang, A.C.Bovik, H.R.Sheikh and E.P.Simoncelli: "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, Vol.13, No.4, pp.600-612, Apr. 2004.
- [3] 板垣秀星, 新井悠祐, 境田慎一, 井口和久, 合志清一, 甲藤二郎: "複数のイントラ符号化形式で圧縮された動画像の画質改善手法の提案", 電子情報通信学会技術報告, IE2008-124, pp.93-98, Dec.2008.