

ユーザ満足度を考慮したシーン制御による高効率動画配信手法

後藤悠斗^{†,a)} 上岡英史^{†,b)}

モバイル端末の普及に伴い、動画を無線ネットワーク環境で視聴する機会が増えてきた。これらの大容量の動画コンテンツはネットワーク帯域を圧迫させる一因であり、それを解消するためにはコンテンツ自体のデータ量をユーザが気にならないレベルで削減することが有効である。本稿では、動画配信時のネットワーク帯域軽減を目的として、ユーザの視覚心理効果を用いた高効率な動画配信手法を提案する。具体的には、コンテンツを細かく区切ったシーン毎に色深度およびフレームレートという動画符号化パラメータ値を変化させ、ネットワーク帯域軽減に最も効果的な値をシーン毎に明確にする。また、それらのパラメータ値から実際にネットワーク帯域軽減度合いを推定することによって提案手法の有効性を議論する。

A Scene-based Efficient Video Content Delivery Considering User Satisfaction

YUTO GOTO[†] EIJI KAMIOKA[†]

As the spread of mobile terminals, there has increased the opportunity to enjoy video contents in wireless network environments. Video streaming data consumes a lot of network bandwidth, therefore the video data size must be reduced without perceiving the content quality deterioration. This paper proposes an efficient video content delivery method using psychological effects in visual perception aiming at reducing network bandwidth. Concretely, the optimized video coding parameters, namely, color depth and frame rate, will be determined for each video scene in the content. In addition, the effectiveness of the proposed method will be discussed by estimating the reduced network bandwidth.

1. はじめに

スマートフォンやタブレット PC 等の処理能力の高いモバイル端末の普及、更には LTE(Long Term Evolution)等高速通信可能なモバイルネットワークインフラの整備に伴い、モバイル端末で高品質動画コンテンツの視聴機会が増えてきている。高品質動画コンテンツを有線 LAN 環境で視聴するのであれば大容量のデータ転送に耐えることができ、ユーザは動画視聴に対してストレスなく、高い満足度を得る。しかしながら、高品質動画コンテンツを 3G や Wi-Fi ネットワークで視聴すると、ネットワーク帯域を圧迫し、他ユーザをも含め、動画視聴に対する満足度を下げるようになってしまう。このような問題を解決するために、動画コンテンツ自体のデータ量を効率よく削減することが必要不可欠である。

また、人間は色情報を常に記憶しておくことができないという色情報への鈍感さや、動きの過程を脳内で補完することができるという、視覚心理効果[1]がある。これらの色情報や動き情報に直接対応する色深度とフレームレートはビデオビットレートを構成する符号化パラメータであり、伝送ビットレート値に大きな影響を与える。

したがって、動画配信時のビデオビットレートを、

この視覚心理に基づいてユーザが気にならないレベルに削減すれば、ネットワーク負荷の軽減となり高効率な動画視聴を期待できる。

また、色深度やフレームレートといった符号化パラメータ値の低下による動画コンテンツの劣化は、コンテンツのジャンルやコンテンツ内のシーンの違いによってユーザが感じる度合いが異なる可能性がある。

本研究ではまず、モバイル環境での動画コンテンツ視聴を想定し、視覚心理に基づいて主観評価実験を行う。具体的には、動画コンテンツ内のそれぞれシーンに対して色深度とフレームレートの変化がユーザ満足度 (QoE: Quality of Experience) [2]に与える影響を観測する。最終的に効率よくビデオビットレートを軽減できる色深度とフレームレートの値をコンテンツ内のシーンごとに明確にすることを目的にしている。

以下、関連研究や筆者らがこれまでに行ってきた先行研究と本研究の違いについて触れ、人間の視覚心理が動画視聴に及ぼす影響についてシーン毎にビデオビットレートを制御する動画配信手法を提案する。また、本提案方式の有効性を確認するために行った主観評価実験について述べ、その実験結果について考察する。

2. 関連研究

色深度やフレームレートを考慮した動画の主観評価実験に関する関連する研究は様々なものがある。John D.

[†] 芝浦工業大学大学院
Graduate School of Engineering and Science, Shibaura Institute of Technology,
3-7-5 Toyosu, Koto-ku, Tokyo, 135-8548 Japan
a) ma11071@shibaura-it.ac.jp
b) kamioka@shibaura-it.ac.jp

Mccarthy 氏ら[2]は動画の色深度の量子化レベルとフレームレートの値を定量的に変化させ、主観評価実験を行っている。これは、単一のコンテンツ内で動画の品質を変化させたとき、ユーザがそのコンテンツを承諾できるかできないかを定めるパラメータを特定するものであり、視聴端末としてデスクトップ PC と PDA の二つを用いている。30 秒の評価時間のうち、どれくらいの時間、そのコンテンツ視聴を我慢できるかを受諾率として評価した。量子化レベルがある一定の値以下になると受諾率は急激に低下するのに対し、フレームレートの値低下には量子化レベルの低下と比較すると、受諾率に及ぼす影響は大きなものではないという結果が得られた。また、デスクトップ PC に対して PDA の方がこの傾向が大きかった。このことから人間はフレームレートよりも量子化レベルによる劣化、つまり、色の劣化に対して敏感であると結論づけた。

この研究では単一のジャンル（サッカーの番組）に対する評価しか行っておらず、他のジャンルに対してはこの実験結果を適応できるかどうかは不明である。これに対し本研究では、コンテンツのジャンルを複数用意し、ジャンルによる評価結果の違いを考察する。

また、F. Agboma 氏ら[3]はシーンではないが一定時間で区切ったセグメント毎に動画の品質を劣化させ、主観評価実験を行っている。これは 3G 携帯端末、PDA、ラップトップコンピュータを用いており、フレームレートとビデオビットレートの値を一定区分の時間毎に劣化させ、その都度被験者に満足できるか否かを問う形式で評価している。そしてサッカーのような動きの激しいジャンルのコンテンツに対しては、被験者は品質の劣化を敏感に感じ取り、ニュースのように動きが少ないコンテンツに対してはあまり劣化を感じづらいつという結果を述べていた。

この研究では、ビデオビットレートの値とは別にフレームレートの値を定量的に変更させているが、ビデオビットレートを構成する符号化パラメータについて深く言及しておらず、更にセグメントも時間単位であり、シーンによって区分していない。本研究では、コンテンツの内容によってシーンを細分化することによって高効率な配信を目指す。

また、先行研究として筆者ら[4][5]はユーザからディスプレイまでの視距離に着目し、視距離を定量的に変化させた際のフレームレートや色深度の変化がユーザ満足度にどのように影響するのかを確認してきた。評価用動画コンテンツは sports, music, animation の 3 ジャンルで、それぞれ 7 コンテンツ分用意し主観評価実験を行った。この研究からシーン毎の実験結果を平均化すると、視距離が遠くなるにつれてユーザは色深度やフレームレートの低下によるコンテンツの劣化に鈍感になる傾向が確認された。図 1, 2 のようにそれによって帯域軽減率が高くなるという結果が得られた。そして、以下のように視距離から帯域軽減率を導く評価式を色深度・フレームレート、そしてジャンル別

に導出した。尚、 x には視距離[cm]を代入する。

・色深度

$$\begin{aligned} \text{sports: Reduced Bandwidth}[\%] &= -4.06 \times 10^{-3}x^2 + 9.56 \times 10^{-1}x - 8.08 \end{aligned}$$

$$\begin{aligned} \text{music: Reduced Bandwidth}[\%] &= -9.90 \times 10^{-4}x^2 + 1.99 \times 10^{-1}x \\ &- 40.08 \end{aligned}$$

$$\begin{aligned} \text{animation: Reduced Bandwidth}[\%] &= -4.36 \times 10^{-3}x^2 + 8.34 \times 10^{-1}x - 7.84 \end{aligned}$$

・フレームレート

$$\begin{aligned} \text{sports: Reduced Bandwidth}[\%] &= -8.17 \times 10^{-3}x^2 + 1.84x - 35.38 \end{aligned}$$

$$\begin{aligned} \text{music: Reduced Bandwidth}[\%] &= -0.59 \times 10^{-3}x^2 + 1.23x - 5.64 \end{aligned}$$

$$\begin{aligned} \text{animation: Reduced Bandwidth}[\%] &= -1.72 \times 10^{-3}x^2 + 4.88 \times 10^{-1}x \\ &- 46.55 \end{aligned}$$

また、ジャンルによって色深度による制御の方が有効なのか、フレームレートによる制御の方が有効なのかは異なるという結果も得られた。また、ユーザが劣化を感じる度合いはコンテンツによって異なるという傾向も得られた。

そこで本稿ではコンテンツを更にシーンによって細分化し、シーン毎に動画を劣化させ主観評価実験を行うことでユーザ満足度にどのように影響が出るのかを明確化する。

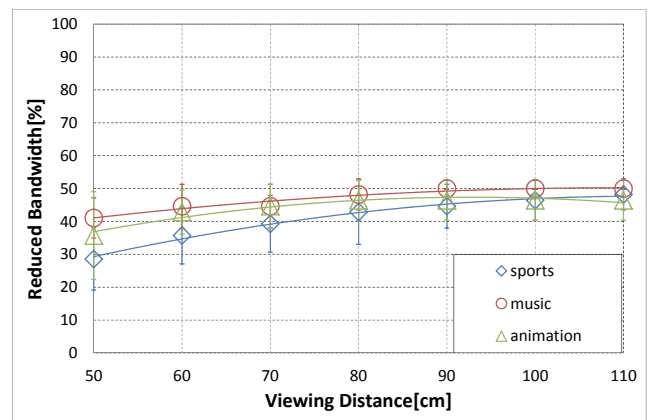


図 1 視距離に応じた帯域軽減率（色深度）

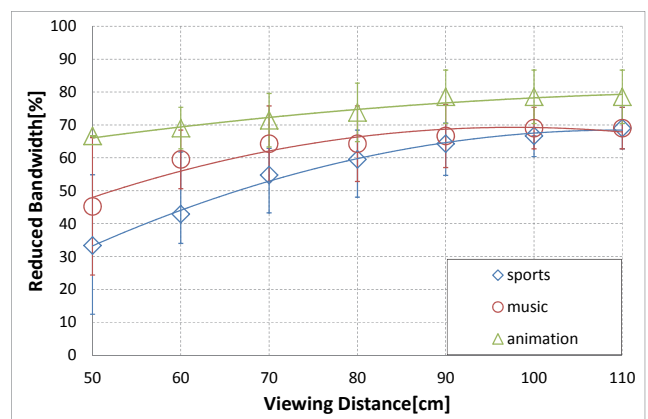


図 2 視距離に応じた帯域軽減率（フレームレート）

3. 提案手法

本研究では、ユーザの視覚心理を利用し、QoEを低下させることなくビデオビットレートを削減する手法を確立する。具体的には、ジャンルやコンテンツを細分化したシーンによって劣化に対するユーザ満足度の傾向が異なると仮定し、シーン毎に色深度やフレームレートを劣化させたコンテンツに対して主観評価実験を行う。そして、それぞれのシーン毎に劣化を感じない最低ラインの色深度値、フレームレート値を明確化することによって、配信における各ジャンル、シーンに応じた符号化パラメータの制御を提案する。

ここで、コンテンツを細分化する際のシーン切り替えの基準について、内容を考慮して以下のように定義する。

- ・ カメラアングルの切り替わり (sports, music)
- ・ 背景画像の切り替わり (animation)

上記のようにコンテンツを細分化したものを主観評価実験で扱うシーンとした。表1に細分化したシーンと各シーンのフレーム数を示す。これを用いることでより詳細な主観評価実験を色深度とフレームレートそれぞれに対して行う。尚、ジャンルの違いによる傾向を確認するために動画コンテンツは sports (サッカーの試合)、music (アーティストのライブ映像)、animation (少女がサーフィンをしている場面) の3ジャンルのものを用いた。

表1 ジャンルとシーン数・フレーム数の関係

	sports				music									
Scene No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Frame数	184	43	130	102	54	62	36	35	52	56	50	33	47	28
	animation													
Scene No.	15	16	17	18	19	20								
Frame数	104	58	52	48	59	52								

主観評価実験では、ビデオビットレートを軽減できる符号化パラメータの明確化が目的であるため、また、フレーム毎に色深度の値を変更する必要があるため、更に、フレームレートをシーン毎に変更する必要があるため、bitmap画像をフレームレートに基づいた枚数分を重ね合わせた無圧縮のAVIフォーマットのものを用いた。

次に色深度とフレームレートの値をそれぞれ変化させるコンテンツの作成手法について述べる。

まず、色深度の評価に用いるコンテンツに関しては、シーン内の全てのフレームの bitmap 画像に関してそれぞれ色深度の値を4~8bppに変更する。作成したコンテンツの符号化情報を表2に示す。

フレームレートの評価に用いるコンテンツに関しては、シーン毎にフレーム数が元の30fpsのものに対して、1/3, 1/2, 2/3, 4/5になるようにフレームの bitmap 画像を意図的に間引き、疑似的に10fps, 15fps, 20fps, 24fpsとなるようにコンテンツを作成する。作成したコンテンツの符号化情

報を表3に示す。

表2 実験用動画画像符号化情報 (色深度)

色深度(bpp)	4	5	6	7	8
解像度	480×242				
ファイルフォーマット	Bit map				

表3 実験用動画画像符号化情報 (フレームレート)

色深度(bpp)	8				
解像度	480×242				
フレームレート(fps)	10	15	20	24	30
ファイルフォーマット	AVI				

尚、今回はフレームレートの傾向を確認する実験においては色深度の値を固定した。また、色深度に関してはシーン内のフレームの内容にあまり差異が見られないため、シーン内の特定の1フレームを用いて静止画として評価する。フレームレートに関しては、先述の手順で作成した動画コンテンツを用いた。尚、今回は画像に評価を集中させるため、音声を含まないコンテンツを使用した。

4. 主観評価実験

ここでは、提案手法の有効性を確立するための主観評価実験手法および実験環境について述べる。

4-1 実験環境

前章にて述べたコンテンツを用いて主観評価実験を行った。被験者にはラップトップコンピュータ (Apple MacBook Air mid 2011, 13.3inch) 上にプレゼンテーションソフトウェアである Keynote[7]を用いて、スライドに静止画や動画画像を埋め込んで図3のように表示した。また、実験手法について記述した教示を実験前に提示し、それに従って被験者にはクリックで自由に次の評価画像に進めるようにした。尚、参照画像は各シーンの最高品質の物 (色深度: 8bpp, フレームレート: 30fps) で、評価画像は順に劣化させていった。また、色深度に対する実験とフレームレートに対する実験ではそれぞれ本学大学院生、学部生、教員を含んだ10人を被験者とした。



図3 主観評価実験用画面イメージ

4-2 実験手法

主観評価実験手順のフローチャートを図4に示す。実験手順は色深度に対してもフレームレートに対しても同様である。尚、評価は sports, music, animation の順に連続で行った。

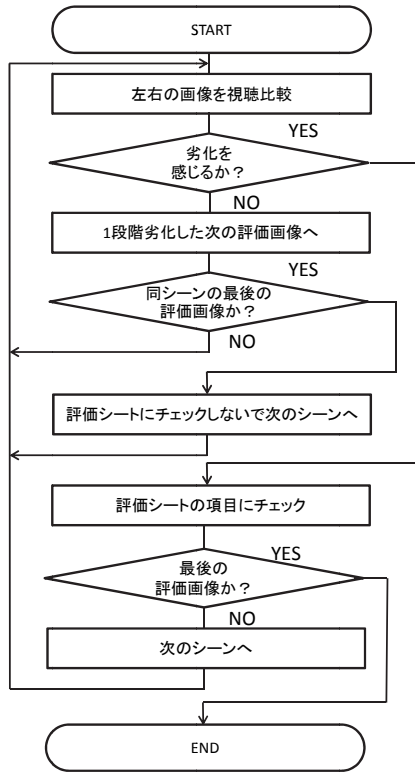


図4 主観評価実験手順フローチャート

5. 主観評価実験結果

ここでは、前章にて述べてきた主観評価実験の結果について述べる。

5-1 色深度の影響

まず、色深度の傾向を確認するために静止画に対して主観評価実験の結果を図5に示す。尚、グラフ中の誤差棒は実験結果の標準偏差、つまりシーン毎の主観評価実験結果のバラつきを意味する。縦軸の適応可能平均色深度は、主観評価実験の際に、「劣化を感じた時点での画像」に 1bpp 加えた値を意味する。すなわち、被験者（ユーザ）が最高品質である 8bpp のものと比較しても劣化を感じないため、その差分のビデオビットレートを軽減できるということである。それはそのまま配信時のネットワーク帯域軽減にも繋がる。つまりこの値が低ければ低いほど高効率な動画像配信が可能となるということである。また、横軸はシーンの番号を表し、棒グラフの色の違いはジャンルの違いを示しており、それぞれ青：sports, 赤：music, 緑：animation である。

色深度に関しては、画像に集中するために静止画を用い

て評価したので、動きに関する考慮はしていない。そんな中、同一ジャンル内でもシーンによって適応可能平均色深度の値にバラつきが生じていることが確認できる。特に、sports はこの傾向が顕著に表れている。評価の高い No.2 のシーンは選手がアップで撮られている場面であり、逆に評価の低い No.3 のシーンはフィールドを引いて撮っているシーンである。アップで寄っている場合、つまり画面サイズに対して選手のユニフォームの鮮やかな色の割合が大きい場合には、色に対する劣化はあまり気にならないと考えられる。反対に引いたシーンでは画面サイズに対して、ユニフォームの鮮やかな色の割合が小さい場合には色に対する劣化を感じ易いと考えられる。

animation も同様で、No.19 の評価が高いのは作中のキャラクターがアップで描かれており、他のシーンと比較すると、背景に対して鮮やかな色の割合が大きいからだと考えられる。しかしながら、music の評価が高い No.6, 9, 12 はカメラが引いたシーンであり、アーティストのアップのシーン等と比べると鮮やかな色の割合が逆に画面に対して少ない。このよう色深度の影響はジャンルやシーンによって異なるということが確認された。

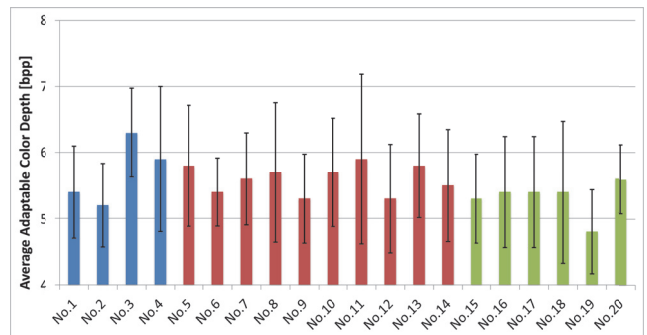


図5 色深度主観評価実験結果（平均適応可能値）

5-2 フレームレートの影響

次に、フレームレートの傾向を確認するために行った主観評価実験結果を図6に示す。グラフの見方は色深度の時と同様であるが、左側の縦軸の適応可能平均フレームレートは実験の際に「劣化を感じた時点での動画」より1段階フレームレートの値を上げることによって被験者が最高品質である 30fps のものと比較しても劣化を感じないということである。右側の縦軸はフレームの数を示しており、各シーンのフレーム数（長さ）を折れ線グラフにて表示したものを追加した。これは、色深度と違いフレームレートの傾向を確認する実験は動画像コンテンツを用いたためである。

まずは sports に関しては、他のジャンルと比較すると評価が低いことが確認できる。sports のコンテンツは動きの激しいシーンで構成されているので、選手やボールの動きに劣化を感じたと考えられる。

また、シーンによって異なるが、全体的にみると他のジ

ジャンルに比べて特に animation のコンテンツへの評価が高いということが確認できる。また No.15 を始め、評価のバラつきも小さいことが確認できた。

music は特に No.9 の評価が高いが、これは舞台照明が明滅するシーンであり、そのタイミングさえ合っていれば被験者は満足できたと考えられる。

次にフレーム数と各シーンの評価の関係についてだが、実験前にフレーム数が少ないシーンはあまりフレームレートの低下によるコンテンツの劣化は気にならないという関係があるのではないかと考えて実験を行った。しかしながら、今回の実験条件下では特にそのような相関関係は見られなかった。

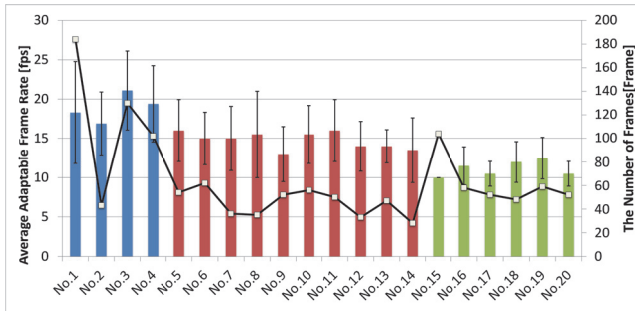


図6 フレームレート主観評価実験結 (平均適応可能値)

つのコンテンツとしたときに適切な値を適応することで高効率な動画配信が期待できる可能性が見出された。

次に、色深度とフレームレートのそれぞれのネットワーク帯域軽減率を比較すると、どのジャンルに関してもフレームレートによる帯域軽減率の方が高いことが確認できる。これは今回の実験環境下では用意した符号化パラメータの段階が色深度とフレームレートで異なることが要因であるが、実際の配信環境を想定してもフレームレートによる制御の方が高い QoE を保てるという点で有効であるといえる。

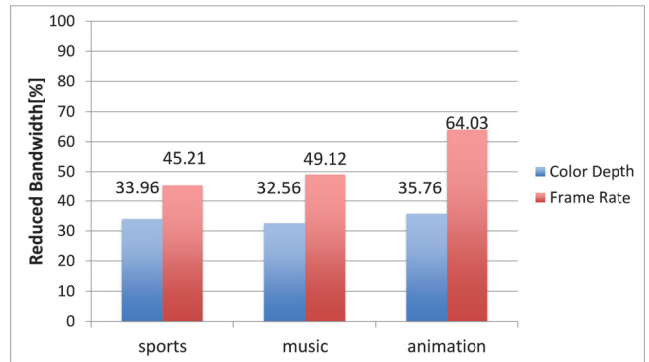


図7 帯域軽減率

6. 考察

これまで色深度とフレームレートに対する主観評価実験結果について、個別にシーン毎に見てきたが、ここではこのシーンを統合して一つのコンテンツとして捉えたときのビデオビットレート削減率、つまりネットワーク帯域軽減率を見る。図7にジャンル別の帯域軽減率を示す。

ここでは、被験者による各シーンに対する評価のうち、もっとも多くチェックされた値(最頻値)を適応可能色深度値・適応可能フレームレート値としてネットワーク帯域軽減率を示している。

色深度による主観評価結果はシーンによってバラつきがあったが、統合して1コンテンツとするとジャンル間あまり差が見られないということが確認された。これにより色深度によるビデオビットレート削減はジャンルによる影響をあまり受けないと考えられる。

フレームレートでの提案手法によるネットワーク帯域軽減率は animation が他の2ジャンルより大きいということが確認できた。つまり animation は特にフレームレートによるビデオビットレート軽減が有効であると考えられる。

また、フレームレートに関しては animation, music, sports の順に高いということが確認できた。先行研究[6]と同様の順で軽減率が高かったが、色深度に関しては傾向が異なるという結果が得られた。尚、同一コンテンツ内でもシーン毎に適応可能色深度値やフレームレート値は異なる。したがって、シーン毎にこれらの値を制御することによって1

7. 結論

表4に各シーンの内容と適応することで高効率な配信が可能になると考えられる符号化パラメータの値をそれぞれ示す。帯域軽減に有効な符号パラメータの値はジャンルやシーンによって異なるということが確認することができた。このようにシーンによって適切な色深度値やフレームレート値を設定することによって、高い満足度を保ちつつ、帯域軽減が可能な配信が期待できる。

表4 シーン内容と適応可能符号化パラメータ値

シーン	ジャンル	内容	適応可能値	
			bpp	fps
1	sports	着距離カメラアングル。バスを回している。	5	15
2		近距離カメラアングル。シュートを打つ。	5	15
3		長距離カメラアングル。ゴールが決まる。	6	24
4		近距離カメラアングル。選手がゴール後のパフォーマンス。	5	10
5	music	近距離カメラアングル。歌手が歌う。	6	20
6		長距離カメラアングル。横から観客全体が映る。	5	15
7		中距離カメラアングル。歌手が歩きながら歌う。	5	15
8		近距離カメラアングル。歌手が歌う。	5	15
9		長距離。ステージ全体が映る。	5	10
10		中距離カメラアングル。歌手が踊り出す。	5	15
11		近距離カメラアングル。歌手が踊る。	6	20
12		中距離カメラアングル。歌手が踊る。	5	15
13		近距離カメラアングル。歌手が踊る。	6	15
14		近距離カメラアングル。歌手が移動する。	6	10
15	animation	少女がサーフィンをしている。徐々に近づいてくる。	5	10
16		引きの画。少女がサーフィンをしている。	5	10
17		寄りの画。少女がサーフィンをしている。	5	10
18		寄りの画。波だけが映る。	5	10
19		寄りの画。少女の顔がアップになる。	5	15
20		別の人が少女を遠目に見ている。	5	10

8. まとめと今後の課題

今回の主観評価実験結果およびそれに対する考察により、シーン毎にユーザが劣化を感じる度合いが異なり、その特性を利用し、シーン毎に色深度の符号化パラメータを変更する制御を行うことによってより細かい単位で精度の高い高効率な動画配信が期待できる可能性が見出せた。

しかしながら、

- ・ 色深度の傾向を見る主観評価実験では静止画を用いており、実際に動画を視聴して評価していない。
- ・ コンテンツ数が各ジャンル1つのみであり、一般化した議論ができない。
- ・ プレイヤーレベルでの実験であり、通信ネットワークを使った動画配信を行っていないので、実際のネットワーク負荷への影響を確認していない。
- ・ 音声が入った時の主観評価への影響を確認していない。
- ・ 現行の圧縮技術との比較をしていない。

といった問題が残っている。また、今回の主観評価実験から得られた適応可能なパラメータ値を用いた動画コンテンツに対して実際に評価を行なうことにより、この提案手法の正当性を確認する必要もある。

また、今回は色深度とフレームレートに対して独立に評価してきたが、今後はこの二つの適応可能なパラメータ値を組み合わせたハイブリッドな動画コンテンツに対して評価をし、それによってQoEにどのような影響をあたえるかを明確にし、それによる帯域軽減率を確認し、議論していく。

尚、シーン数やジャンルから適応可能な色深度値・フレームレート値を導出する定式化についても議論し、QoEを低下させず、伝送ビットレート軽減するという点で、高効率動画配信を行う提案手法の有効性を定量的に評価していく。

参考文献

- 1) 内山恵二監修, 映像情報メディア学会編, 視覚心理入門-基礎から応用感覚まで-, オーム社, 2009年
- 2) 高橋玲, “マルチメディア QoE 評価技術の標準化動向,” 電子情報通信学会技術研究報告, Vol. 107, No.19 CS2007-5, CQ2007-9, April 16, 2007, pp. 49-54.
- 3) John D. Mccarthy, M. Angela Sasse, Dimitrios Miras, “Sharp or Smooth? Comparing the Effects of Quantization vs. Frame Rate for Streamed Video,” Proceedings of the SIGCHI conference on Human factors in computing systems, p.535-542, April 24-29, 2004, Vienna, Austria.
- 4) F. Agboma, A. Liotta, “User centric assessment of mobile contents delivery,” in 4th International Conferences on Advances in Mobile Computing and Multimedia, Yogyakarta, Indonesia, 2006, pp. 121-130.
- 5) 後藤悠斗, 上岡英史, “視覚心理と符号化情報を用いた高効率動画配信手法,” 電子情報通信学会技術研究報告, Vol.111, No.384, MoMuC2011-40, January 19, 2012, pp. 19-24.
- 6) 後藤悠斗, 上岡英史, “視覚心理を用いた高効率動画配信手

法,” 日本人間工学会関東支部大会講演論文集, B1-5-2, December 10-11, 2011, pp. 52-53.

7) アップル - Keynote- <http://www.apple.com/jp/iwork/keynote/>