

視聴者表情から動画ダイジェストを作成するシステム

古川 裕士^{1,a)} 濱川 礼²

概要: 本論文では、ユーザの表情を撮影した動画 (以下、表情動画とする) を用いて、ユーザの嗜好を反映させ個々のユーザに適した動画ダイジェストに反映させるシステムについて述べる。本システムは動画を一度視聴した後、改めて自分が気に入ったシーンだけを楽しむ動画ダイジェストを作成するために使用する。ユーザは本システムのプレイヤー部分で動画を再生し、同時に Web カメラでの撮影が開始される。動画の再生終了後、Web カメラにより撮影した顔動画像を解析し、ユーザの表情が大きく変化した部分を動画から抽出し、動画ダイジェストを自動作成する。本システムの動画ダイジェストならば仮定ではなくユーザが実際に盛り上がり、表情が変化したシーンをそのまま動画ダイジェストにできる。本システムを用いて実際に動画を視聴し、動画ダイジェストを作成・視聴までの評価を 24 名の評価者に対し実施した。結果のアンケートでは、映像中の対象物から動画ダイジェストを作成する手法と比べて高評価を得ることができた。

The video summary production system that uses viewer's facial expression

FURUKAWA HIROSHI^{1,a)} HAMAKAWA REI²

Abstract: This system uses the facial expression of a viewer, and produces a video summary automatically. Currently, there are many methods of producing a video summary automatically. This system takes a face of user who watch a video by WEB camera. We define a taken video as "the face video". This system analyzes the face video by image processing, and consequently, recognizes the facial expressions. This system produces a video summary based on the facial expressions. 24 students who belong to Hamakawa laboratory of the Chukyo University evaluate this system.

1. はじめに

本論文では、ユーザの表情を撮影した動画 (以下、顔動画像とする) を用いて、ユーザの嗜好を反映させ個々のユーザに適した動画ダイジェストに反映させるシステムについて述べる。ここで、動画ダイジェストとは動画コンテンツの要約を指す。

本システムは動画を一度視聴した後、改めて自分が気に入ったシーンだけを楽しむ動画ダイジェストを作成するために使用する。現在、世の中には動画ダイジェストを自動作成する研究やシステムが数多くあり、多くの動画ダイジェスト作成手法は、映像中の対象物や映像の音声を解析することで、その変化を抽出し、動画ダイジェストに反映

している。これらの手法には、問題点が二つある。まず、ほとんどの場合、解析の効果を高めるために「ジャンルを限定している」ということが挙げられる。確かに例えばサッカーに限定することで、ゴールシーンの検出等の解析精度が高まりより良い動画ダイジェストに成りうる。しかし、多様な動画がインターネット他に存在する現在ではさまざまなジャンルの動画ダイジェストを作成するために、それぞれ別の動画ダイジェスト作成システムを使用しなければならない。次に「ユーザの嗜好に関わらず同一動画ダイジェストを多数のユーザが利用する」ことを暗黙に仮定していることである。我々はダイジェスト動画に出来るだけユーザ個人の嗜好を反映させたいと考えた。

我々はこれらの問題を解決するために視聴中の顔を撮影し、顔動画像を用いることで、どの部分でユーザが盛り上がったかを判定し、動画ダイジェストに反映するシステム

¹ 中京大学大学院 情報科学研究科 情報科学専攻

² 中京大学 情報理工学部

^{a)} honmaruyori@yahoo.co.jp

を開発した。

本システムは、1台のPCと1台のWebカメラを使用する。Webカメラをユーザの顔が映るようPCに取り付ける。ユーザは本システムのプレイヤー部分で動画を再生し、同時にWebカメラでの撮影が開始される。動画の再生終了後、Webカメラにより撮影した顔動画を解析し、ユーザの表情が大きく変化した部分を動画から抽出し、動画ダイジェストを自動作成する。視聴中の顔を撮影し、その表情動画を用いることで、どの部分でユーザが盛り上がったかを判定し、動画ダイジェストに反映するシステムを実装し、検証を行った。

2. 背景

2.1 動画ダイジェスト自動作成の現状

近年、世の中には動画ダイジェストを自動作成する研究やシステムが数多くある。また、動画ダイジェスト自動作成機能を搭載したブルーレイディスクレコーダーも販売されており、代表としてSONYのBDZ-RX105[1]の動画ダイジェスト再生機能や三菱のDVR-BZ340[2]の見どころ再生などが挙げられる。このことから動画ダイジェスト自動作成が一般的に用いられていることが分かる。

2.2 関連研究・システム

多くの動画ダイジェスト作成手法は、映像中の対象物や映像の音声を解析することで盛り上がっている部分を判定している。例えば[3]ではゴールポストが動画中に存在している部分をゴールシーンつまり重要なシーンと位置づけ、動画ダイジェストを作成する。さらに[4]に代表されるように映像のキーフレームシーンから一つ特徴的な画像(キャプチャー)を抽出し、類似度を測り、関連したシーンを検索する手法もある。また[5]ではマイクによって録音された会議中の発言を音声認識し、発話の多い部分を動画ダイジェストに反映する。映像中ではないが、関連したtweetからアメリカンフットボールの試合の自動ハイライトを作成する[6]や動画再生中のコメント数によって、動画ダイジェストを作成する[7]もある。さらに野球中継のテロップの情報エントロピーに基づいて重要度の高いシーンを動画ダイジェストとして切り出す[8]もある。また近年では動画を切り出して編集するのではなく、無駄なシーンを高速再生する研究[9]も進められている。

[10]は本システムと同様の手法の研究である。視聴者の表情をウェブカメラで観測して[11]を引用しつつ感情分析をしている。顔の動きを3次元的に測定し、感情パラメータを通してダイジェスト動画を作成する

2.3 現在の動画ダイジェスト自動生成の問題点

現在の動画ダイジェスト自動作成の問題点をあげる。

- ジャンルを限定している

本システムは前述の関連研究と比べ、「動画のジャンルが限定されている」という問題点を解消した。ユーザの視聴動画を解析で使用するため、バラエティ動画・映画・ニュース動画などでも動画ダイジェストを作成できる手法を提案している。関連研究であげた[3]、[6]、[8]は解析の効果を高めるためにそれぞれサッカー動画、アメリカンフットボール動画、野球中継限定でしか動画ダイジェストを作成しない。このようなスポーツ動画からダイジェスト動画を作成する場合は、スポーツのジャンルを限定する研究が多くある。[5]は会議中の状況を撮影し、使用することを目的としており、[7]は動画のジャンルは限定されていないが、ニコニコ動画のコメントが解析に必要なため、ニコニコ動画にアップロードされている動画しか動画ダイジェストを作成できない。

これらは全て動画中の特徴物を解析し、動画ダイジェストに反映しているのに対し、本システムでは視聴している動画は一切考慮せず、視聴者の顔動画のみを解析対象としている。

- ユーザの嗜好に関わらず同一動画ダイジェストを多数のユーザが利用すると仮定している

多くの自動動画ダイジェスト作成手法は、同一ダイジェスト動画を多数のユーザが利用する仮定で動画ダイジェストが作成されている。例えば[7]では単位時間あたりのコメント数が多いシーンほど、多数のユーザが盛り上がっていると推測し、動画ダイジェストを作成している。このような手法では、コメント数は少ないが、ユーザ自身が盛り上がっている部分は動画ダイジェストに反映されない。また、逆にコメント数は多くても、個々のユーザが盛り上がっていない部分も動画ダイジェストに反映されてしまう。実際に、同研究の評価アンケートでは、「面白い部分が見られなかった」や「動画ダイジェストの内容が物足りない」という回答があった。[8]はテロップの情報エントロピーの大きい出来事(発生しにくい事象)を特筆すべき重要なシーンと仮定し、その部分では多くのユーザが盛り上がるかと推測し、さらに動画の音声を解析し、動画ダイジェストを作成している。この手法では、テロップではルール上劇的な状況ではないが、視聴者自身は盛り上がっているシーンが取得できない。さらには野球中継の見方を珍プレーのみ視聴するとした場合、この手法では笑いの動画ダイジェストは作成できない。

3. 提案手法

本システムで作成する動画ダイジェストは、一度視聴した動画を盛り上がるシーンだけに短縮して再度視聴することを目的としている。ユーザは見たいと思った動画を本システムで再生し視聴する。見終わった後にもう一度見たいシーンだけを抽出し動画ダイジェストを作成する。

動画ダイジェストをユーザが視聴する理由は大きく分けて三つある。一つ目は「動画ダイジェストを本編の代わりに視聴する」。関連研究で代表例をあげると [3] が当てはまる。どちらも試合としても視聴するのではなく、劇的なシュートシーンや点が動いたシーンなどを集めた動画ダイジェストを作成する。[4]、[6] もこれに当てはまる。二つ目は「本編を見る前の参考として動画ダイジェストを視聴する」.[7] が代表例で、映像コンテンツが爆発的に増加する近年では、閲覧や選別の時間を短縮するための参考に動画ダイジェストを作成している。最後に「本編を視聴した後、もう一度見たいシーンを動画ダイジェストで視聴する」.[5] が代表例で会議での話し合いを経て、その後重要な話し合いのシーンだけを抽出し、無駄な会話シーンをカットする。これを後で視聴することでメモすることなく、話し合いの内容を集約できる。

本システムは「本編を視聴した後、もう一度見たいシーンを動画ダイジェストで視聴する」を目的として動画ダイジェストを作成している。[10] も本システムと同様に、「本編を視聴した後、もう一度見たいシーンを動画ダイジェストで視聴する」ことを目的としている。しかし、[10] は評価時のみ特徴量をグラフとして表示しているが、本システムでは特徴量のグラフがユーザにとって動画ダイジェストのジャンル選択に有効と考え、動画編集時常に表示している。また、[10] は表情の変化が顕著なシーンを動画ダイジェストとしているが、本システムは表情変化の少し前のシーンを動画ダイジェストとすることで、よりユーザの感情変化にこだわった。

さらに、本システムは関連研究と比べ、「動画のジャンルが限定されている」という問題点を解消した。ユーザの視聴動画さえあれば、あらゆるジャンルの動画で動画ダイジェストを作成できる手法を提案している。関連研究では動画中の特徴物を解析し、動画ダイジェストに反映しているのに対し、本システムでは視聴している動画は一切解析・考慮することなく、視聴者の顔動画像のみを解析対象としている。

また、「同一動画ダイジェストを多数のユーザが利用すると仮定している」という問題点を解消する方法として、ユーザの嗜好を素直に動画ダイジェストに反映させるために、ユーザの表情を用いることにした。現在、販売されているパソコンには Web カメラが標準で搭載されているた

め、動画視聴中の表情を撮影することが容易になってきている。動画視聴中のユーザの表情は、動画の内容によって変化する。面白いシーンでは笑い、興奮するシーンでは、力が入った表情になる。動画ダイジェスト作成に表情を用いることでユーザの感情、つまり映像中でユーザの心が動いたシーンを判定することができると考えた。つまり、「表情変化が多いシーンは重要なシーンである」と仮定する。そして、ユーザの表情ごとに作成した動画ダイジェストの抽出されたシーンが違えば、一般化した動画ダイジェストよりも、より個人の嗜好にあった動画ダイジェストを作成できる。

学生 24 人を対象にアンケートを実施した。「動画の視聴中に表情が変化していると思うか」の問いに対して、YES は 20、NO は 4 と回答を得た。このアンケート結果から、大人数の人が動画の視聴中に表情が変化していることがわかる。

機能としては、動画を再生すると同時にユーザの顔の撮影を開始する。動画再生終了後、Web カメラにより撮影した顔動画像を解析する。切り出した各部位から特徴点を抽出し、表情の変化量(以下、表情量とする)を算出する。表情量を基にユーザの表情が大きく変化した部分を動画から抽出し、動画ダイジェストを自動作成する。

図 1 に本システムの全体構成を示す。

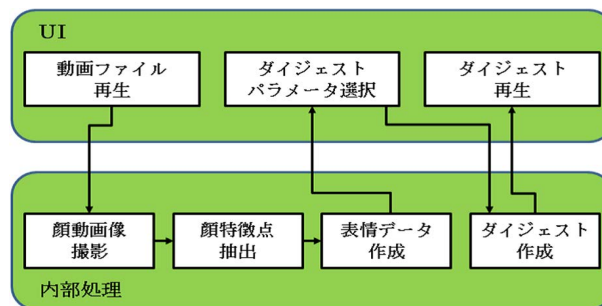


図 1 システムの全体構成図

再生直後 25 フレームですぐ山場を向かえる動画は少ないので撮影開始から 25 フレーム間は盛り上がる部分がなく、動画視聴中の顔を動画として保存した際、ユーザの顔が変化しなかった。そこで、撮影開始から 25 フレーム間の特徴点の平均座標を表情量を算出する際に基準にする座標とする。さらに、特徴点算出方法には、[12] の手法を用いる。これは各フレームの特徴点座標と基準座標の特徴点間の距離と三角形領域の面積を使用する手法で、顔全体のゆがみも考慮できる。

本システムでは、実際に動画を視聴し、Web カメラで撮影した顔動画像から表情の観察を行い、長い表情でも 50 フレーム以上は続いていたことから、50 フレーム毎に平均値を算出している。

4. 機能詳細

4.1 顔特徴点抽出

顔特徴点抽出部では、画像処理を用いて、2値化画像を作成し、眉・目・口から表情認識に用いる特徴点を抽出する。

顔画像内での顔部位の位置関係は、人やカメラ位置によってほとんど変わらない。そのため、予め決定した領域を使用して眉、目、口の部位がある領域を切り出す。毎秒5フレームの顔動画を使用し、OpenCV[13]の学習データを用いて、顔領域を切り出す。切り出した顔領域の画像を図2に示す。



図2 顔領域の切り出し画像

まず、眉領域から輝度値の小さい部分を抽出し、2値化画像を作成し、横幅の最も大きな領域から左右の端の点を取得する。次に、目領域から彩度の小さい部分（白目部分）とエッジ抽出した輪郭の2値化画像を作成し、上下左右の端の点を取得する。そして、口領域から輝度値の小さい部分と唇の色を抽出し、2値化画像を作成し、上下左右の端の点を取得する。作成した2値化画の例をそれぞれ図3に示す。左の画像から眉毛/目/口の2値化画像となっている。

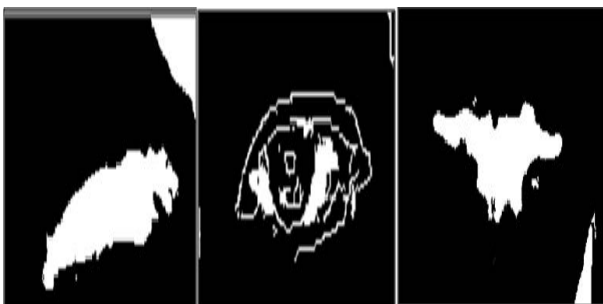


図3 2値化画像

4.2 表情データ作成

盛り上がりを判定する表情は、Ekmanらによって提案された基本6感情[11]を用いた。動画ダイジェスト作成に必要な感情である怒り・嫌悪・悲しみを除いた笑い・

興奮・恐怖の3種類とし、表情量の算出はそれぞれに対応した特徴点を選出する。例えば、笑いは口角が上がり目尻が下がるため、基準座標と現在の表情の座標の目の特徴点と口の特徴点の距離を表情量の算出に用いる。

顔特徴点抽出部で抽出された特徴点座標の平均を算出する。算出された座標をユーザの基準の表情とし、各フレームとの変化量を求め、表情量として定義する。

表情量の算出は、ユーザの顔の大きさや顔の傾きなどによる表情量への影響を低減するために[12]で用いられている以下の2つの方法を用いる。

- 直線距離—基準座標の2点間の距離と各フレームの2点間の距離の差を求める。

基準座標の2点を (x_1, y_1) と (x_2, y_2) とする。あるフレーム L の座標を (a_1, b_1) と (a_2, b_2) とするとあるフレームの表情量 F_L は

$$F_L = \sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2} - \sqrt{(a_2 - a_1)^2 - (b_2 - b_1)^2}$$

と表せる。

- 三角形面積—基準の座標の3点を結んだ三角形の面積と各フレームの3点間の距離の差を求める。

基準座標の3点を $X(x_1, y_1)$ と $Y(x_2, y_2)$ と $Z(x_3, y_3)$ とする。あるフレーム L の座標を $A(a_1, b_1)$ と $B(a_2, b_2)$ と $C(a_3, b_3)$ とするとあるフレームの表情量 D_L は

$$D_L = \frac{(x_1 - x_2)(y_3 - y_2) - (y_1 - y_2)(x_3 - x_2)}{2} - \frac{(a_1 - a_2)(b_3 - b_2) - (b_1 - b_2)(a_3 - a_2)}{2}$$

と表せる。

図4に特徴点の直線距離、図5に三角形面積の計算例を示す。

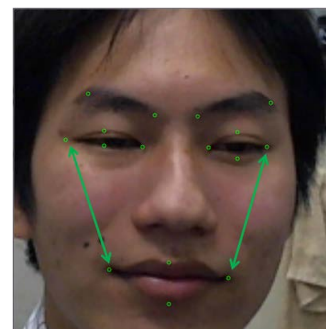


図4 特徴点の直線距離計算

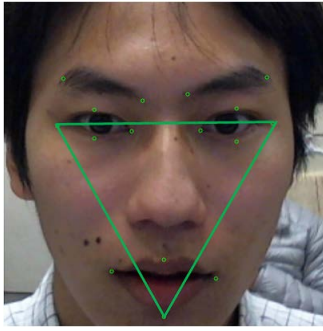


図 5 特徴点の三角形面積計算

笑いの表情量を F_H (初期値は 0) とし、基準座標の目の特徴点と口の特徴点の距離を F_T 、現在の表情の座標で目の特徴点と口の特徴点の距離を F_1 とすると、

$$F_H = F_H + |F_T - F_1|$$

のように表せ、表情量に加算していく。

さらに、笑いは目が細くなるので、基準座標と現在の表情の座標の目の上端と下端の距離を表情量の算出に用いる。笑いの表情量を F_H (先ほどの算出結果) とし、基準座標の目の上端の特徴点と下端の特徴点の距離を F_E 、現在の表情の座標で目の上端の特徴点と下端の特徴点の距離を F_2 とすると、

$$F_H = F_H + |F_E - F_2|$$

のように表せ、先ほどの算出結果に加算していく。興奮や恐怖も同様に追加していく。例えば、興奮は顔全体が外側に広がるので、眉の特徴点と口の特徴点の三角形面積を表情量の算出し、表情量に追加する。恐怖は眉間をしめこめる。つまり、左右の眉の特徴点の距離が短くなるので、左右の眉の特徴点の距離の差を、表情量に追加する。

フレーム毎に算出した表情量を、横軸にフレーム数、縦軸に表情量をもつグラフに表示する。そのグラフをダイジェストパラメータと呼ぶ。ダイジェストパラメータは、横軸に時間、縦軸に表情量として構成される。図 6 にダイジェストパラメータの例を示す。

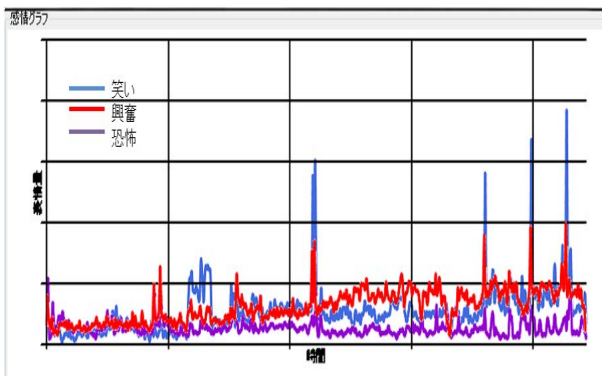


図 6 ダイジェストパラメータ

ダイジェストパラメータは笑い・興奮・恐怖の 3 種類作成される。青/赤/紫のグラフはそれぞれ笑い/興奮/恐怖のダイジェストパラメータを表す。ユーザは 3 種類の中から、動画ダイジェストを作成したい感情のダイジェストパラメータを選択する。

4.3 動画ダイジェスト作成

選択されたダイジェストパラメータを用いて、動画ダイジェストを作成する。

まず、試作段階で動画と視聴者の表情変化の関係を調査したところ、視聴者の表情はおおよそ 50 フレーム単位で変化することが判明したので、ダイジェストパラメータの値から各フレームの直近 50 フレーム毎に移動平均値を算出し、移動平均線を求める。移動平均線とは、一定時間の数値の平均値を算出しグラフ化したものである。本研究では、平均値を算出しグラフ化を行い、移動平均線とする (図 7)。

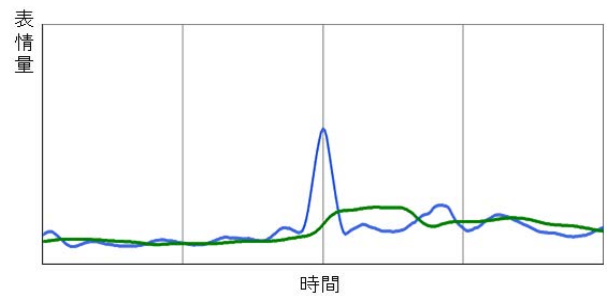


図 7 移動平均線

次に、ダイジェストパラメータと移動平均線を用いて動画ダイジェストに用いるシーンを選出する。ダイジェストパラメータの中で移動平均線が表情量を上回った場合にユーザが盛り上がっていると判断を行い、図 8 の水色の部分をシーン候補とする。

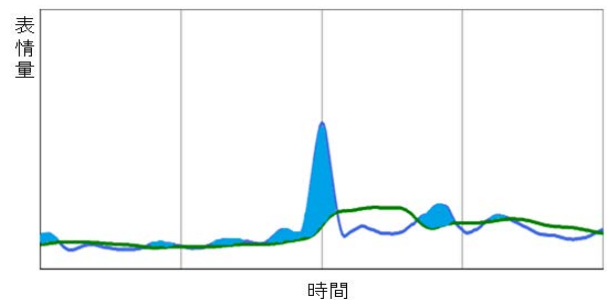


図 8 シーン候補

そして、シーン候補となった水色の部分の面積を算出していき、移動平均線以上となった感情量と平均値との差を

求める。求めた差が縦軸、時間が横軸となり面積を算出していき、算出した面積の大きい順から再生開始位置、再生終了位置と割り当てていく。再生位置の決定は、ダイジェストパラメータの表情量が移動平均線を上回った位置を再生開始位置とする。下回った場合は再生終了時間とする。

ただし、笑いのダイジェストは純粋に笑っているシーンのみを抽出することは、極端に言えば種明かしの部分だけをつなげる作業であり、それが適切な要約として機能するかは不明である。そこで開発過程のダイジェスト作成のテストから、笑っているシーンよりも前からその笑いのための布石が置かれるため、再生開始位置を3秒前に移動した。その再生開始位置と再生終了位置の指定を繰り返し行うことで動画ダイジェストを作成する。

ここで、実際に作成したシステムのUIと右下に本来、本システムのUIには表示していない動画ダイジェスト作成時の顔とダイジェストパラメータを示す(図9)。



図9 UIとユーザの顔・ダイジェストパラメータ

5. 評価

本システムの目的であるユーザが盛り上がった部分が動画ダイジェストに反映しているかを確認するために、研究室所属の学生24人を対象に評価を行った。

評価は15分の動画(バラエティ動画とスポーツ動画、映画、ドキュメンタリー)を視聴してもらい、それぞれ3分の動画ダイジェストを作成し、自分の表情から作成した動画ダイジェストと他人の表情から作成した動画ダイジェストについてアンケートを実施した。本システムは、笑い・興奮・恐怖の3種類の動画ダイジェストを作成できる。用意したバラエティ動画とスポーツ動画で一番視聴者の表情が出やすい笑いと興奮、恐怖のダイジェスト動画を評価した。

比較として映像の特徴から動画ダイジェストを作成する研究[14]を使用し、本システムで作成した動画ダイジェストと比較してもらった。

さらに、動画を視聴する間にもう一度みたいと感じたシーンまたは感情が動いたシーンを書き出してもらい、そのシーンが動画ダイジェストに反映されているかを確認した。本システムが重視している「感情が変化しているシーンが動画ダイジェストに反映されているか」という問いに、図10のような回答を得た。

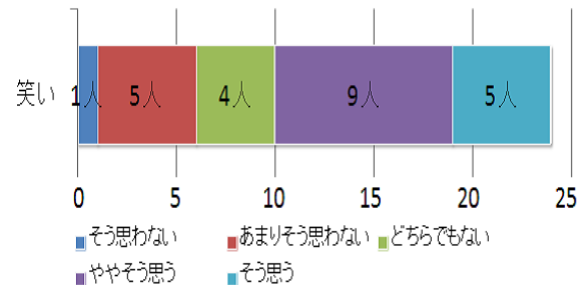


図10 評価結果1

本システムで作成した動画ダイジェストと[14]で提案されている手法で作成した動画ダイジェストのどちらがもう一度みたいシーンが含まれていたかの問いに、93%が本システムで作成した動画ダイジェストの方が含まれていたと答えた。

その結果を図11に示す。

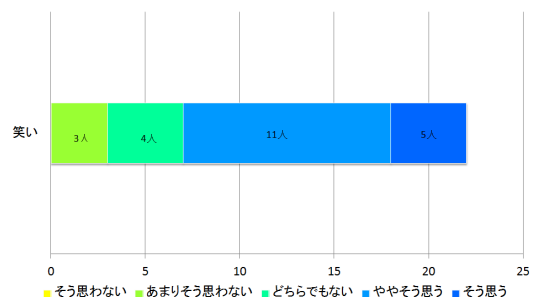


図11 評価結果2

本システムで作成した笑いのダイジェストパラメータと視聴者がもう一度みたいと思ったシーンとを比較した。これが一致していれば、映像中にユーザの心が動いたシーンを、本システムが正確に抽出していることを意味する。視聴者が選んだシーンを赤い枠線で囲み、比較例を図12に示す。

青いグラフである笑いのダイジェストパラメータの表情量が極端に増加する部分のほとんどが、赤い枠線が囲まれているのが分かる。比較によると評価者がもう一度みたいシーンを概ね動画ダイジェストに反映できていると判断できる。同様に興奮、恐怖のダイジェストパラメータから作成したスポーツ動画、映画、プロフェッショナルの動画ダ

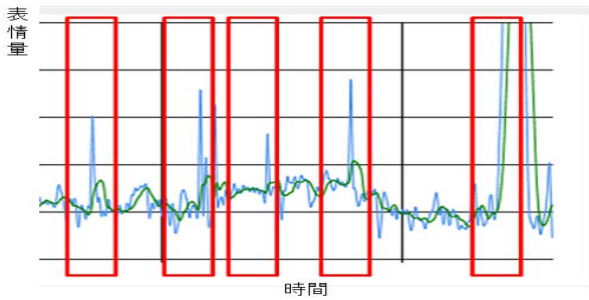


図 12 評価結果 3

イジェストも評価者がもう一度みたいシーンを動画ダイジェストに反映することができた。

6. 考察

同じ動画を視聴したある評価者 A, B, C, D のダイジェストパラメータを図 13, 14, 15, 16 に示す。

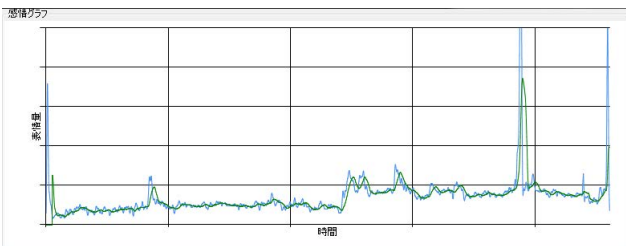


図 13 評価者 A のダイジェストパラメータ

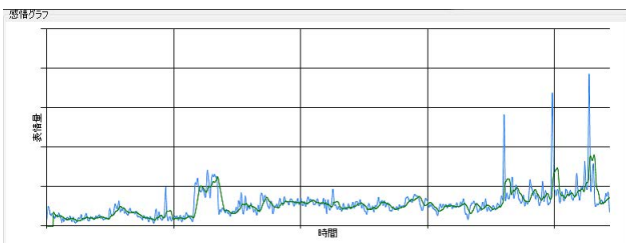


図 14 評価者 B のダイジェストパラメータ

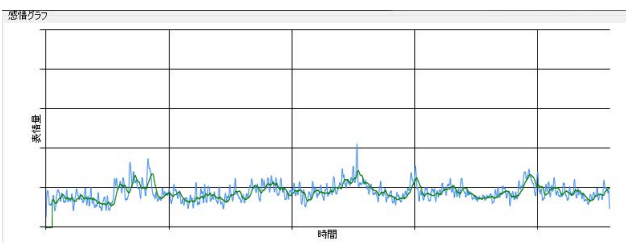


図 15 評価者 C のダイジェストパラメータ

評価者 A, B ではダイジェストパラメータの右側に大きな盛り上がりがあることが分かる。盛り上がりの程度や若干のシーン場所は異なるが、バラエティ動画の笑いのツボや興奮するスポーツのシーンなどは個人によって大きく異なることはない。つまり、別の人または複数人のダイ

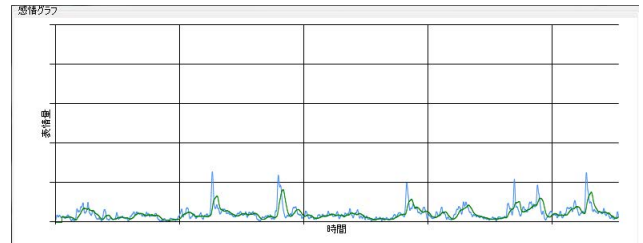


図 16 評価者 D のダイジェストパラメータ

ジェストパラメータを使用することで、ある程度汎用性が期待でき、多くの人が共感できる一般化された動画ダイジェストを作ることができる。

しかし、評価者 A, B とは違いダイジェストパラメータの左側に大きな盛り上がりはなく、評価者 C にはむしろ前半と中盤に、評価者 D は全体的に盛り上がりを感じていない。同じ動画でも、視聴者によって別のシーンに盛り上がりを感じていたり、少人数でも盛り上がる場合があることが確認できる。このことから本システムを使用することで、一般化された動画ダイジェストだけでなく、数人が盛り上がりを感じる複数ユーザ専用クラスターリングされた動画ダイジェストを作成することも可能である。

今回評価に使用したバラエティ動画とスポーツ動画、映画、ドキュメンタリーでは、評価結果からユーザがもう一度みたいと思ったシーンを十分反映した動画ダイジェストが作成できた。さらに、映像から動画ダイジェストを作成するよりも、ユーザの好みにあったシーンを反映することができることを確認した。特にお笑いの動画から、笑いの動画ダイジェストを作成する際に半数以上から高評価を得られた。

7. まとめ

評価と成果から、映像の特徴や音声に着目する手法では取得できない視聴者のもう一度見たいと思ったシーンを反映した動画ダイジェストが作成できた。ダイジェストパラメータとユーザの表情とが正確に対応しているかを確認するため、評価で動画を視聴中のユーザの顔と視聴時の笑いのダイジェストパラメータを表示し、図 17 に示す。上がダイジェストパラメータ、下がユーザの顔である。

図より、笑いのダイジェストパラメータが上昇するときに、笑い始めているのが分かる。さらに、評価時のアンケートで次のような意見もあった。

- 感情のグラフを見ながら元動画を再生してみたい
- 個人で使うには動画の解析に時間がかかりすぎる

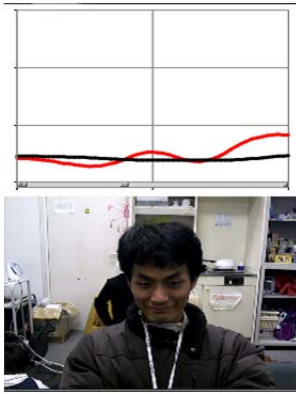


図 17 ダイジェストパラメータと表情

上記の「感情のグラフを見ながら元動画を再生してみたい」を受けて、動画ダイジェスト視聴時にも視聴中のユーザの顔とその時のダイジェストパラメータを表示するように改良した。

今回のシステムの実装と評価から動画のジャンルによっては、「どのような表情に基づいて動画ダイジェストを作るべきなのか」という点を追求することが極めて重要な課題であることが分かった。そのため、動画ジャンル別のどのような表情になるかを学習させていく必要がある。

さらに、評価アンケートにもあったように「個人で使うには動画の解析に時間がかかりすぎる」という問題点がある。本システムは毎秒5フレームの顔動画を使用しているので作業量が多く、15分の動画から3分の動画ダイジェストを作成するのに約2時間かかる。スポーツ動画や映画は大抵2時間の動画時間があるので、それを視聴し、3分のダイジェスト動画作成する場合にかかる処理時間は約8時間となる。現在、1秒5フレームずつ画像処理しているので、1秒間に処理するフレーム数を減らせば、速度は上がるが、システムとしての精度は下がってしまうという問題点がある。つまり、ユーザがもう一度見たいシーンを手動で編集したほうが、手間はかかるものの短い時間で作業が終了できてしまうため、提案システムの存在意義が問われてしまう。

8. 今後の課題

今後、視聴者の表情解析をトレーニングする手法や3Dモデリングを使用する表情解析手法を導入し、学習した表情とマッチングするような機能を追加する。さらに、学習データを使用することで解析時間を短縮するとともに、動画要約以外の方向性に提案アプローチを応用することにより汎用性を高め、挽回したい。

Webカメラが標準で付属したパソコンが増えているので、このシステムは動画ダイジェストを作成するためのものだけでなく、日常的なパソコンの機能(電子書籍やインターネットサーフィン)をダイジェストとしてまとめることが

できると考える。具体的には、成績評価の要素とするために学習履歴情報から学習者の行った学習のダイジェストを作成する手法を提案した[15]のように動画ではないコンテンツをダイジェストにする研究も数多くある。そこで、パソコンの画面遷移をダイジェストにする機能の追加し、パソコン上で作業するプロセス、例えばイラストやブログを見ている表情を使用し、それらのコンテンツをダイジェストにするなどのようにシステムの汎用性を高める。

以上の構想を[16]などで、社会に提案していきたい。

参考文献

- [1] SONY BDZ-RX105 入手先 (<http://www.sony.jp/bd/products/BDZ-RX105/>)
- [2] 三菱 DVR-BZ340 入手先 (<http://www.mitsubishi-shielectric.co.jp/bd/products/dvrbz340/index.html>)
- [3] 山本大樹他 サッカー映像のシーン自動解析の研究 電子情報通信学会研究報告 vol.104 no.537 (2005)
- [4] ZeeshanRasheed et.al "Detection and Representation of Scenes in Videos" IEEE TRANSACTION ON MULTIMEDIA vol.7 no.6 (2005)
- [5] 富田章裕他 思考状態と発話停止点を利用した会議の動画ダイジェスト生成支援情報処理学会論文誌 vol.47 (2006)
- [6] Anthony Tang et.al "EpicPlay : crowd-sourcing sports video highlights" CHI2012 (2012)
- [7] 小川一昭他 視聴者からのコメント情報を用いたダイジェスト動画疑似生成の提案 情報処理学会研究報告 vol.33 (2009)
- [8] 片岡充照他 情報エントロピーに基づく野球中継番組のダイジェスト自動生成 映像情報メディア学会誌 (2010)
- [9] 栗原一貴 動画の極限的な高速鑑賞のためのシステムの開発と評価 WISS2011 (2011)
- [10] Joho et.al "Exploiting Facial Expressions for Affective Video Summarization" ACM (2009)
- [11] P.Ekman 表情分析入門 誠信書房出版 (1987) 工藤力訳
- [12] 顔特徴点を用いた特徴選択と特徴抽出による表情認識に基づく映像中の表情表出シーン検出 DEIM Forum (2011) C1-5
- [13] OpenCV 入手先 (<http://opencv.jp/>)
- [14] 伊藤秀和 濱川礼 限られた視聴時間内における動画の効果的な時間短縮手法 電子情報通信学会技術研究報告 vol.108 (2009)
- [15] 李正遠他 学習履歴情報に基づいた学習過程のダイジェスト化 vol.103 (2003)
- [16] 古川裕士 濱川礼他 視聴者の表情を用いた動画ダイジェスト作成システム 情報処理学会全国大会講演論文集 vol.74 (2012)