

映像視聴者の映像再生位置操作と眼球運動に 基づく映像要約手法

澤田和弥[†] 吉高淳夫[†]

個人で大量の映像を管理することが容易となり、映像内容の振り返りや重要な場面の再確認を効率化することが望まれている。この問題を解決する手段の1つとして、個人にとって重要な場面のみを残した要約映像を生成することが考えられる。本研究では個人の嗜好に沿う要約映像を生成するため、映像視聴時の視聴者の映像再生位置操作と眼球運動の情報から視聴者の注目場面を検出する。映像再生位置操作とは、映像視聴時の早送り操作や巻き戻し操作などである。映像再生位置操作は常に表れる反応ではないため、映像再生位置操作で取得できない視聴者の注目場面は、眼球運動から取得する。眼球運動は、視聴者の映像に対する反応が常に表れ、特に停留頻度は、視聴者が場面を注目して視聴しているか否かを判定する一指標になる。評価実験の結果から、視聴者の映像再生位置操作と停留頻度を合わせることで、個々の注目場面を検出することが可能となり、また個人の嗜好に沿う要約映像を生成できる可能性を示した。

Video Summarization Based on Viewers Operating Video Playback and Eye-movement

Kazuya Sawada[†] and Atsuo Yoshitaka[†]

Since inexpensive large-capacity data storage devices have become available, storing videos has become much easier. In order to recall the content of the video, watching it from the beginning is not always efficient. One of the solutions for it is to summarize the video for the purpose of shortening the viewing time. As a result, only the important scenes from the video are extracted. The authors proposed a video summarization method for individual viewers. In order to generate a video summary which fits the preference of individual viewers, the proposed method utilizes two types of user information: the viewer's eye movement and video playback operation while watching video. The video playback operation is defined as fast-forward and rewind operation when watching the video. The proposed system estimates the desired scene by using video playback operations. In addition, it also analyzes eye movements to obtain the desired scene which was not captured by the video playback, since eye movements represent the viewer's reaction to the video. In particular, fixation frequency is used in the proposed system since fixation frequency indicates whether the viewer is paying attention to the scene. The system generates the resultant video summary by detecting

two types of scenes: the scenes which were revisited by the viewer using playback operations, and the scenes which are accompanied by a certain fixation frequency. The evaluation indicated that the system was able to detect important scenes for individual viewers and to make the video summary for individual viewers.

1. はじめに

映像のデータ量はテレビ放送、インターネット配信などの普及に伴い増加の一途をたどっている。また、大容量かつ安価なデータ記憶装置(例:HDD等)が普及し、個人で大量の映像を保存、管理することが容易となった。これにより、TV放送を映像録画機器(例:HDDレコーダー等)で録画後に視聴することや、日常の風景等をビデオカメラで撮影し、その映像を保存することも一般的となった。個人が管理する映像は、内容の振り返りや重要な場面の再確認を目的に再視聴されることがある。その際に、最初から映像を視聴する行為は、目的以外の場面に視聴時間が割かれるため効率的ではない。この問題を解決する手法の一つに、映像内の重要な場面のみを残した要約映像を生成する映像要約手法がある。

映像要約手法の研究は、映像の種類や目的とする要約場面の性質によって様々なアプローチがある[1]。そのアプローチは大きく分けて2つある。1つ目は映像内の特徴、情報のみから重要場面を検出する映像内容解析手法である。この手法は、1つの入力映像に対して1通りの要約映像が生成されるため、個人の嗜好に沿った要約映像を生成することは難しい。2つ目は映像視聴時の視聴者の情報から重要場面を検出する視聴者情報解析手法である。1つの入力映像に対して個々の視聴者の要約映像が生成されるため、個人の嗜好に沿う要約映像を生成することが可能である。個人が保存、管理する映像を対象に要約映像を生成する場合、その個人が映像を1度は視聴する可能性は高い。また、再視聴される映像である場合は興味を持ち視聴している可能性も高く、個人の嗜好に沿う要約映像が望まれると考えられる。そのため、このような映像に対しては視聴者情報解析手法を用いるのが有効であると考えられる。

個人の重要場面は様々であり、どのような視聴目的を持つかによって変化する。例えば、サッカーなどのスポーツ映像をエンターテインメントとして視聴する場合、ゴールシーンや歓声の起こったシーンなどの視聴者の情動(驚きや笑い等)を喚起する場面が重要場面となる。一方、サッカーの上達を目的に視聴する等、映像から詳細な情報を得る目的で視聴する場合、特定選手のドリブルやパス回しなど技術的な情報が含まれている場面が重要場面となる。視聴者の情動を喚起する場面を検出するためには瞳孔径[2]や顔表情などの視聴者の情報が有効であるが、情動を喚起する場面が少ない映像に対しては有効ではない。スポーツの技術向上のための講座映像や語学勉強のための

[†]北陸先端科学技術大学院大学 情報科学研究科
School of Information Science, Japan Advanced Institute of Science and Technology

映像等、情動を喚起する場面が少ない映像においても個人の嗜好に沿う要約映像が望まれるだろう。

本研究は、個人が映像の詳細な情報を得る目的で視聴する映像を対象に、個人の嗜好に沿う要約映像を生成することを目指す。本研究においての重要場面は、視聴者が目的意識を持ち、映像内容から詳細な情報を得ようと注目する場面である。また、要約映像の定義をこの注目場面を連結した映像とする。

詳細な情報を得ようと視聴する映像に対しては、驚きや悲しみなどの情動を喚起させる場面は殆ど無く、詳細な情報を得ようと注目した場面の視聴者の反応を得る必要がある。映像に対する視聴者の反応には、意識的な反応と無意識的な反応がある。意識的な反応の1つに映像視聴時の行動がある。意識的な反応は、視聴者自らの意志により起こるため映像に対する反応が明確である。常に視聴者は行動を起こすとは限らないが、特に重要な場面を検出できると考える。一方、無意識的な反応の1つに眼球運動がある。眼球運動は映像に対する反応が常に表れるため、多くの注目場面を検出できる可能性がある。代表的な眼球運動に注視と追従があり、注視は対象物の詳細な情報を得るときに発生し、追従は対象物を眼球の中心に捉えるときに発生する。視聴者の意識的な反応と無意識的な反応として本研究が着目したのは、映像の再生位置操作と停留頻度である。映像再生位置操作とは、録画映像の早送りや巻き戻しなどの操作である。早送り操作時は見る必要がない場面を飛ばし、巻き戻し操作時は再び見たい場面にアクセスする目的があると考えられる。そのため映像に対する視聴者の意図を推測できる可能性がある[3][4]。また停留頻度は、絵画鑑賞時において停留時間の長い状態が密に発生するときは「特定の探索」を行なうという知見があり[5]、この反応から視聴者の注目場面の検出が可能であると考えられる。これらの映像視聴時の視聴者の情報を用いて、本研究では要約映像の生成を目指す。

2. 関連研究

本研究の手法と同じ映像視聴時の視聴者の情報を基に要約映像を生成する視聴者情報解析手法に関する研究について述べる。視聴者の生体信号である視点座標、瞳孔径、心拍を用いて、映像内容解析手法に個人の嗜好を加え要約映像を生成する研究がある[6]。しかしながら、対象とする映像ジャンルが映画に限定されている。また、アニメ、映画、スポーツ等の映像を対象に視聴者の瞳孔径から要約映像を生成した研究があるが、重要場面の検出率が約40%と高い値は示せていない[7]。個人で撮影した映像を対象とし、映像視聴時のサッカード(跳躍運動)、瞬目、頭部の動き、顔表情から要約映像を生成した研究がある[8]。対象とする個人で撮影した映像は、結婚式などエンターテインメント要素が高い映像であり、本研究が対象とするスポーツの技術向上

の為等の映像には対応していない。このように、個人の嗜好に沿う映像要約の関連研究は、映画やドラマ、コメディ等、視聴者の情動を喚起する場面が多い映像を対象としている。そのため、情動場面で反応が表れやすい瞳孔径などの生体信号や、映像視聴時の顔表情を用いており、本研究が対象とする映像においては他の視聴者の情報に着目する必要があると考える。

3. 提案手法

3.1 映像再生位置操作

映像再生位置操作とは、録画映像視聴時などに映像の再生位置を操作することである。具体例では、ビデオプレーヤーで操作する早送り操作や巻き戻しなどがある。映像再生位置操作は、一般的な視聴環境において普及しており、視聴者の意志により映像の再生位置を自由に調整できることから、映像に対する意図が大きく表れる行動である。操作時の意図として、早送り操作は見る必要がない場面を飛ばし、巻き戻し操作は再び見たい場面にアクセスする等の目的があると考えられる。本研究では、多くのビデオプレーヤーに搭載されている一般的な映像再生位置操作を操作可能とした。その操作と一般的な機能を表3.1に示す。表3.1の映像再生位置操作に対応する視聴者の意図を推定するために予備実験を行った。予備実験の内容は、映像再生位置操作を被験者の意志で行える状態で映像を視聴してもらい、映像視聴後に各映像再生位置操作の意図に関するアンケートに記入してもらうである。アンケートの質問は、「〇〇(早送りなど)操作は、どのような意図で使いましたか？使用しなかった場合は、使用する場合の意図をご記入ください。」とした。

表 3.1 映像再生位置操作の一覧

操作名	一般的な機能
再生	通常再生を行う
早送り	再生速度を早める
巻き戻し	再生位置を戻す
一時停止	再生を止める
コマ送り/コマ戻し	1フレーム進む/戻る、再生を止める

表 3.2 操作意図のアンケート結果のまとめ

操作名	操作意図	回答割合
再生	他の操作をキャンセル	8/9
早送り	目的の場面への移動	8/9
巻き戻し	見たい場面を再び見る	9/9
一時停止	コマ送り/コマ戻しと同じ意図	7/9
コマ送り/コマ戻し	詳細な動きの確認	8/9

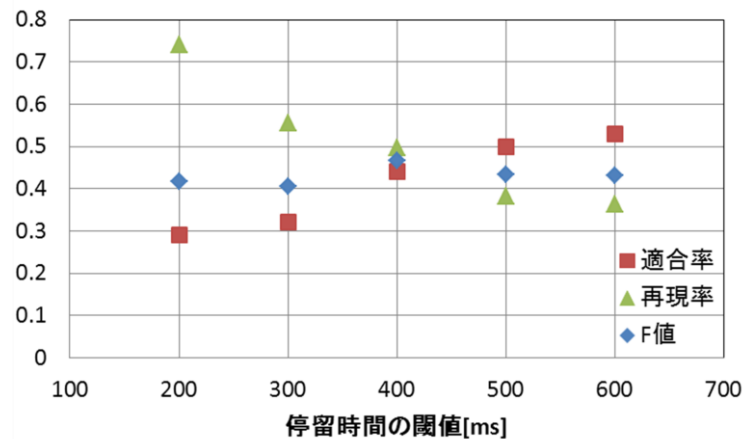


図 3.1 停留時間の閾値を変化させた際の適合率, 再現率, F 値

被験者 9 名のアンケート結果を表 3.2 にまとめる。表 3.2 の結果から、多くの視聴者は見たい場面を再び見るや詳細な動きを確認するために、「巻き戻し操作」、「一時停止」、「コマ送り/コマ戻し」を行うとわかる。そこで提案手法では、「巻き戻し操作」、「一時停止」、「コマ送り/コマ戻し」で再び視聴された場面を視聴者の注目場面とする。

3.2 停留頻度

視線がある一定の範囲内に留まり注視することを停留という。停留と判定する範囲の条件はいくつか提案されているが、本研究では半径 50 pixel 以内の視線を停留と定義する。視線の停留を用いて注目場面を検出する従来研究として絵画鑑賞時には

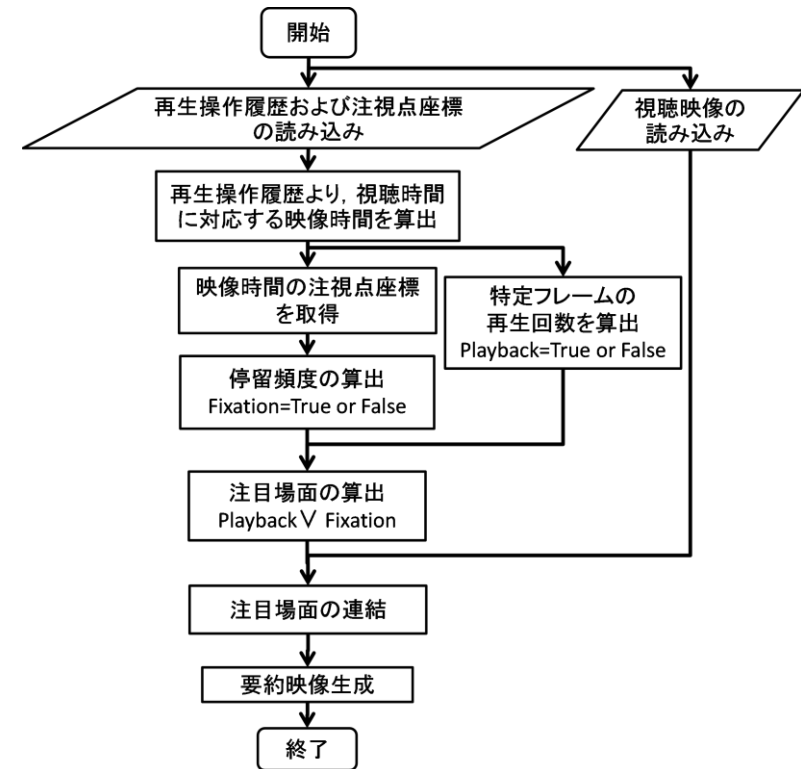


図 3.2 映像要約生成プロセス

300ms 以上 3000ms 未満の停留を 3 回以上連続で発生した時を注目状態と判別できるという知見がある[5]。この知見は、静止画における知見であるため、動画像に適するようにサッカーの試合映像の実験データを用いて閾値を決定した。停留時間の閾値を変化させたときの、適合率, 再現率, F 値の結果を図 3.1 に示す。図 3.1 の結果から、F 値が最も高いときの停留時間 400 ms を閾値と決めた。停留が連続して発生する回数は、参考文献[5]と同じ 3 回以上とした。よって、提案手法では 400 ms 以上の停留が 3 回以上発生した場面を視聴者の注目場面とする。

3.3 映像要約生成プロセス

提案手法の映像要約生成プロセスを図 3.2 に示す。まず、映像視聴時の視聴者から得られた映像再生位置操作の再生操作履歴および注視点座標を読み込む。読み込まれ

た再生操作履歴から、視聴時間(絶対時間)に対応する映像時間(相対時間)を算出する。映像時間の長さは、視聴する映像の長さに対応している。映像時間からフレームの再生回数を算出し、2 回以上再生されたフレームを注目場面とする。また、注視点座標から停留頻度を求め、3.2 で定めた条件を満たしたフレームを注目場面とする。映像再生位置操作と停留頻度より得られた注目場面の論理和を要約映像候補の注目場面とする。注目場面を連結する際に、短すぎる場面は要約映像において不快となるため削除する。削除する条件は、「注目場面の長さが 40 フレーム(約 1.3 秒)以下」かつ「前後の注目場面の間隔が 40 フレーム以上」とした。40 フレームは、停留頻度の条件より得られる最小時間から決定した。

3.4 実装した視聴者の情報取得システム

実装した環境の概要を図 3.3 に示す。映像視聴時の視聴者の映像再生位置操作と停留頻度を取得するシステムについて述べる。本実験では、市販されている HDD レコーダーの機能をリモコン信号受信装置(mbed, 赤外線センサで実装)と動画プレイヤー(GOMPLAERY)にて再現することで、映像再生位置操作の操作履歴自動で取得できるようにした。また、ディスプレイ一体型視線追従装置 Tobii T60 は視聴映像を表示しながら視聴者の眼球運動を測定する。デスクトップパソコン HP Compaq DC7900 は、視聴映像の再生と映像再生位置操作の再生履歴および注視点座標の保存を行った。

4. 評価実験

4.1 視聴映像, 被験者

詳細な情報を得る目的を持ち視聴する映像として、今回視聴映像として選んだのは NHK で放送されたサッカーの試合(インテル v.s. ミラノ)[9]である。熱心に映像を視聴してもらうため、被験者が興味を持ち視聴できる映像を選んだ。TV 放送では、ゴールシーンのリプレイやシュート場面のリプレイがあるが、それらは映像編集者が視聴者に対して重要な場면을提示する行為であるため削除した。また、映像時間は 5 分間とした。また、被験者人数は大学院生 11 名であり、全ての被験者はサッカーに対して興味がある。

4.2 実験手順

実験は以下の手順(1)から(5)の順で行った。

手順(1): 実験内容の説明: 被験者が実験室の照明や雰囲気慣れるまでの間、実験内容の説明を行った。被験者に対して視聴映像の簡単な説明、映像の長さ、視聴指示、実験の終了条件、映像再生位置操作が利用可能なことを説明した。視聴指示内容は、「サッカーの上達を目的に、視聴後映像は二度と見ることが

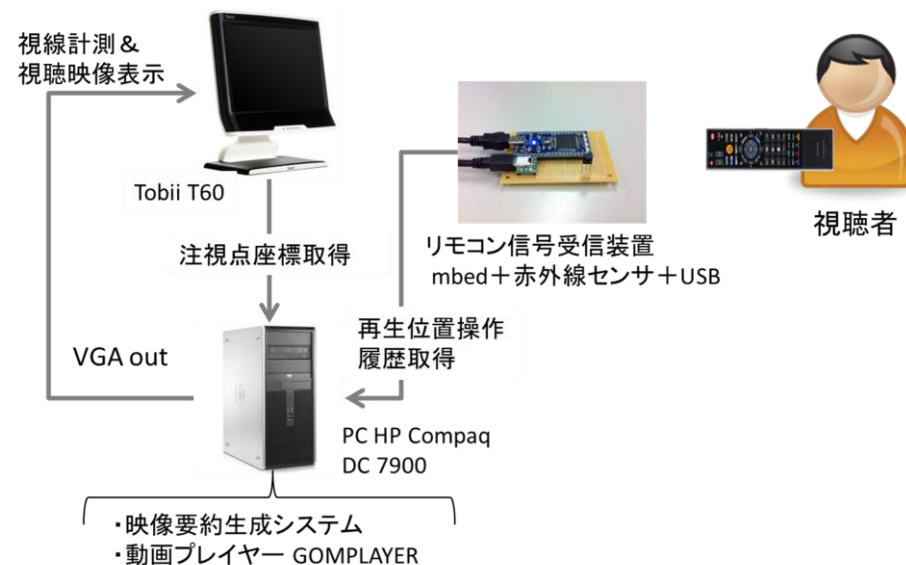


図 3.3 視聴者の情報取得システム

できないつもりで熱心に視聴して下さい」である。

- 手順(2): 映像再生位置操作の練習: 視聴映像以外の映像で「早送り」、「巻き戻し」、「一時停止」、「コマ送り/コマ戻し」操作を確認した。操作に慣れたと被験者が判断した時点で映像再生位置操作の練習を終了した。
- 手順(3): 視線検出キャリブレーション: 映像視聴時の視線のキャリブレーションは、Tobii T60 を用いて検出した。キャリブレーション後に映像の視聴を開始した。
- 手順(4): 映像視聴の開始&終了: 被験者が再生操作を行った時点で映像の視聴を開始し、映像を最後まで見終わった時点で映像の視聴を終了した。
- 手順(5): アンケート記入: 映像視聴終了後、被験者はフレーム番号が表示された映像を再び視聴し、注目した場面のフレーム区間を記入した。また、選んだ各注目場面に対して 3 段階の重要度(高, 中, 低)を記入してもらった。

4.3 客観評価-適合率, 再現率, F 値-

提案手法が検出する注目場面の精度評価として、適合率, 再現率, F 値を以下の式より求めた。本研究において、適合率が高いほど、映像再生位置操作時は注目場面である割合が高く、再現率が高いほど、注目場面で映像再生位置操作を行う割合が高い

ことを表す. Ground Truth は, 実験手順(5)で各被験者が記入したフレーム区間とした. Detected frame は, 提案手法から算出されたフレーム区間である.

$$P = D \cap G / D \quad (1)$$

$$R = D \cap G / G \quad (2)$$

$$F = 2PR / (P + R) \quad (3)$$

P(Precision) : 適合率, R(Recall) : 再現率, F(F-measure) : F 値,
D(Detected frame) : 提案手法により注目場面であると検出されたフレーム,
G(Ground Truth) : 被験者が明示した注目場面のフレーム,
 $D \cap G$: 被験者の明示した注目場面のフレーム番号と提案手法の注目場面のフレーム番号が一致したフレーム

4.4 主観評価

提案手法が生成した要約映像を, 被験者 11 名に評価してもらった. 主観評価の方法は, 各要約映像に対して以下の 2 つの質問に対する評価を 5 段階で記入してもらう方法である. 5 が最も高い評価, 1 が最も低い評価とした. この主観評価は, 評価実験終了後, 3 週間以内に実施した.

質問 1 : 「あなたが注目したい場面が含まれている映像ですか？」

質問 2 : 「要約映像として満足した映像ですか？」

質問 1 は, 客観評価で評価した注目場面の検出を主観的に評価するための質問である. 質問 2 は, 生成された要約映像の全体としての評価を行うための質問である. また, 被験者が評価する要約映像は以下の 3 つである.

- ・提案手法 (Our method) : 評価する被験者から生成された要約映像
- ・比較手法 1 (Constant) : 一定間隔で場面をランダムに抽出し生成された要約映像
- ・比較手法 2 (Other) : 他の被験者から生成された要約映像

それぞれの要約映像の長さは, 提案手法の要約映像とほぼ等しい映像とした. 比較手法 1 は, 他の映像要約研究においてベースライン手法として採用されている. 無造作に場面を抽出した要約映像よりも優れていることを評価できる. 比較手法 2 は, 評価する被験者以外から生成された要約映像である. 提案手法が個人の嗜好に沿う要約映像を生成できたかを評価できる.

4.5 各情報で得られる注目場面の重要度の割合

映像再生位置操作で得られた注目場面と停留頻度で得られた注目場面の重要度の割合を観察するため, 各重要度の注目場面の再現率を以下の式より求めた.

$$R' = I / D' \quad (4)$$

R' : 各重要度の再現率, I : 各重要度の総フレーム数(被験者 11 名分の総和),
 D' : 各重要度で検出された総フレーム数(被験者 11 名分の総和)

4.6 実験結果

被験者 11 名(A~K)の適合率, 再現率, F 値の平均値を図 4.1, 図 4.2, 図 4.3 に示す. 図 4.1 には停留頻度のみを用いた結果, 図 4.2 には映像再生位置操作のみを用いた結果, 図 4.3 には停留頻度と映像再生位置操作の論理和をとった提案手法の結果を示す. 図 4.4 には被験者 11 名の主観評価の結果の平均, 図 4.5 に各重要度の注目場面の再現率を示す.

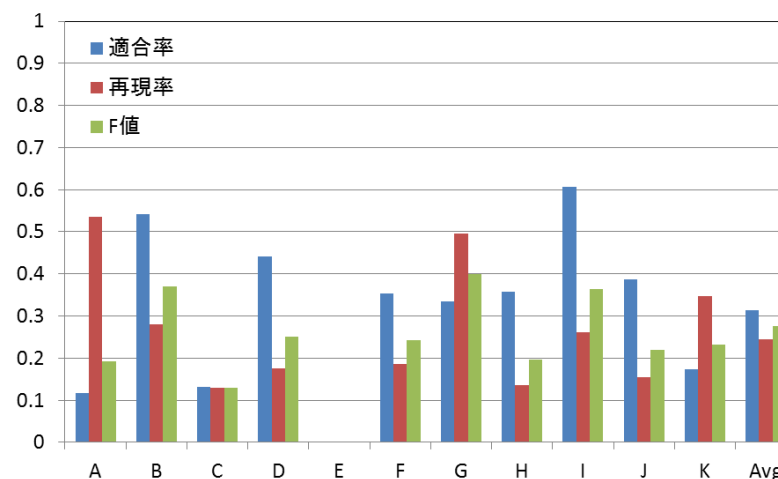


図 4.1 停留頻度による適合率, 再現率, F 値

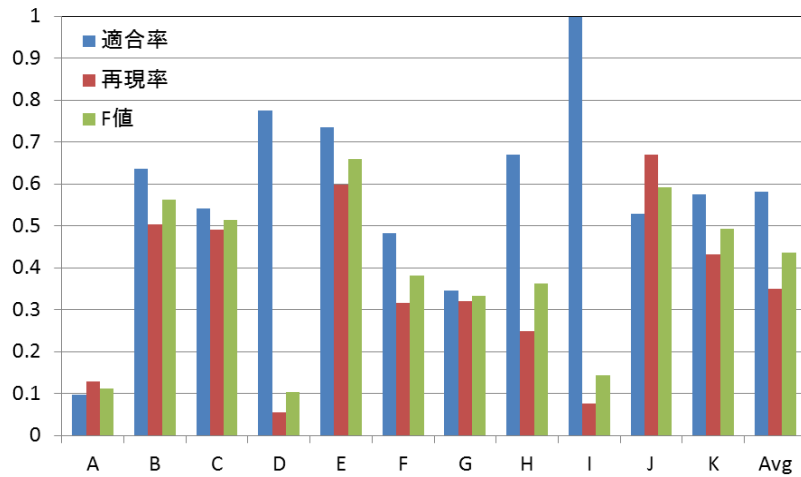


図 4.2 映像再生位置操作による適合率，再現率，F 値

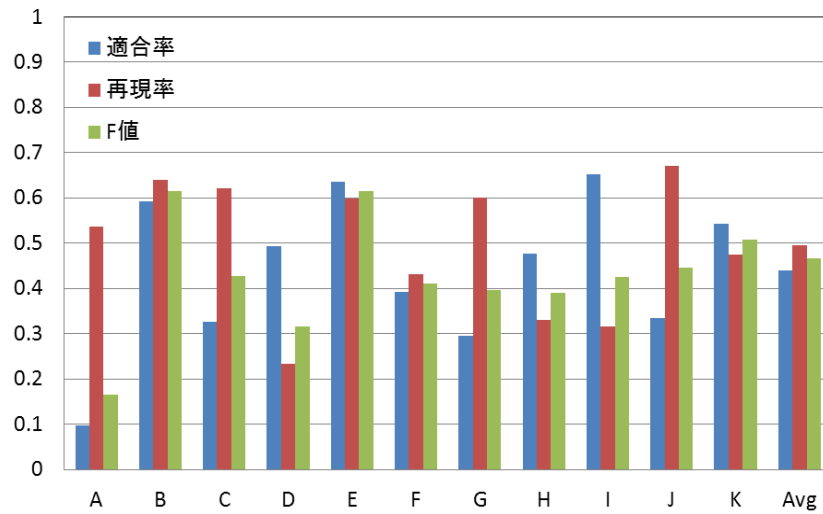


図 4.3 停留頻度と映像再生位置操作による適合率，再現率，F 値

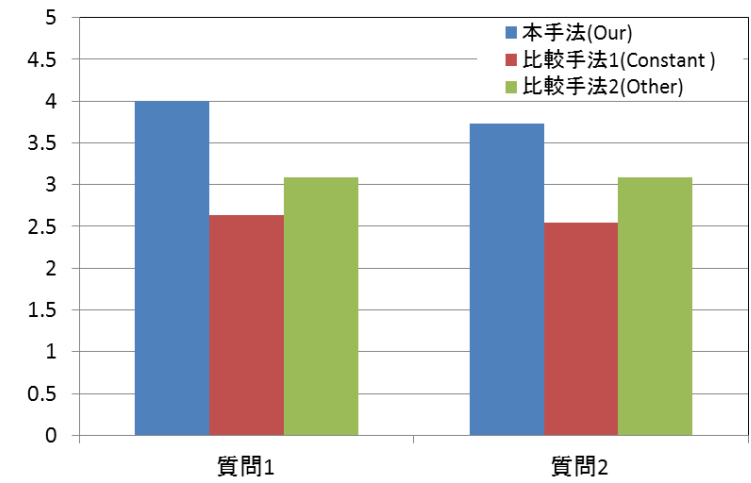


図 4.4 主観評価の結果

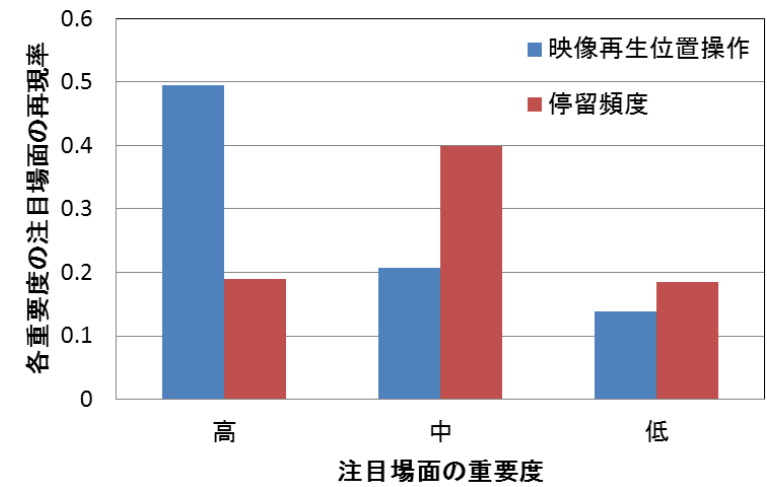


図 4.5 各重要度の注目場面の再現率

4.7 考察

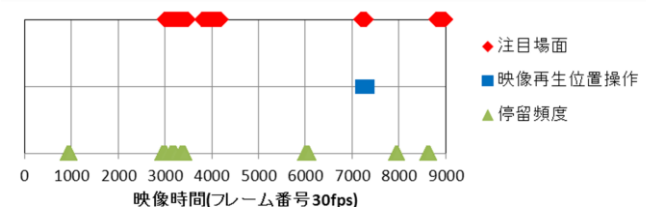
図 4.1 の結果から、停留頻度の適合率、再現率共に高い値ではないことがわかる。しかしながら、映像再生位置操作で検出できなかった場面の検出が行えている可能性がある。また、停留頻度による注目場面の検出条件を、停留時間のみ検討したが停留の連続発生回数は参考文献[5]の条件と同じであるため、今後検討する必要があると考える。被験者 E は、実験の途中で測定範囲から外れてしまい、視線を計測出来なかったため注目場面を検出する事が出来なかった。図 4.2 の結果から、映像再生位置操作は全体として再現率に比べ適合率は高いことがわかり、本研究が対象とする映像と視聴状態において注目場面の検出に有効だと示唆される。被験者 A は、今回用意した視聴映像に対して実験後のアンケートにて興味を持てなかったと回答しており、本研究の対象とする視聴条件を満たせなかった。被験者 I は、視聴映像においてリプレイがあると思っており、実験内容が適切に伝わっていなかった。

図 4.3 の結果から、停留頻度と映像再生位置操作を合わせることで F 値が向上したのがわかる。停留頻度の F 値は 0.28、映像再生位置操作の F 値は 0.44、両方を合わせたときの F 値は 0.47 である。これは映像再生位置操作で検出できなかった注目場面を停留頻度で検出できている可能性があることを示している。この結果の分析として、図 4.6 にランダムでランダム選んだ被験者 4 名の映像時間に対する注目場面、映像再生位置操作、停留頻度を載せた。この結果から、停留頻度のみを用いて検出されている注目場面があることがわかる。

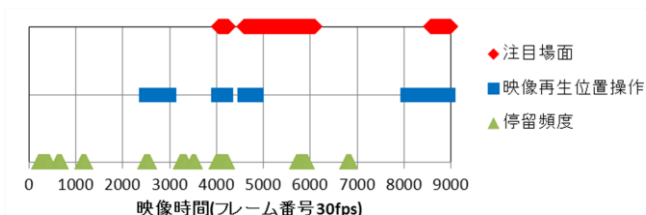
図 4.4 の結果から、質問 1 と質問 2 の両方において、提案手法の要約映像は他の比較手法の要約映像よりも高い評価を得ていることがわかる。このことから、提案手法により個人の嗜好に沿う要約映像を生成できると考えられる。図 4.5 の結果から、重要度の高い注目場面における再現率は映像再生位置操作が最も高いとわかる。よって、映像再生位置操作は特に重要な注目場面を検出できると考えられる。映像再生位置操作は常に起こる視聴者の反応ではないが、特に重要な注目場面の検出に有効であると考えられる。また、停留頻度は重要度が中の注目場面を多く検出できていることから、視聴者が無意識に注目している場面を検出できており、注目場面検出の有用な情報であると考えられる。

5. まとめ

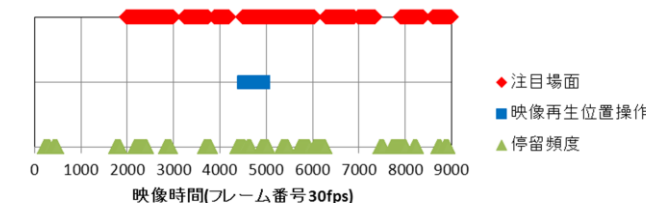
本研究は、個人が映像の詳細な情報を得る目的で視聴する映像を対象に、個人の嗜好に沿う要約映像の生成を行った。映像視聴時の視聴者の意識的な反応である映像再生位置操作の検出と無意識的な反応である視聴者の停留頻度の計測を行い、視聴者が注目した場面の検出を行った。映像視聴時における映像再生位置の各操作に対する意図を被験者 9 名に対して調査したところ、視聴者は再び見たい場面や詳細な動きを確



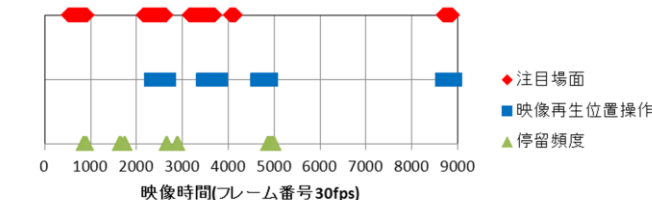
(a) 被験者 D



(b) 被験者 F



(c) 被験者 I



(d) 被験者 K

図 4.6 各被験者の映像時間に対する注目場面、映像再生位置操作、停留頻度

認するために、「巻き戻し」、「一時停止」、「コマ送り/コマ戻し」操作を行うことがわかった。また、静止画において視聴者の停留頻度から注目場面の検出が可能であるという知見を、本研究が対象とする動画像に適応するため、停留時間の条件を検討した。提案手法は、映像再生位置操作により再視聴された場面と停留頻度の注目状態を判別する条件を満たした場면을視聴者の注目場面として検出した。提案手法の評価実験を、被験者 11 名に対しサッカーの試合映像を対象に行った。提案手法により検出した注目場面の精度を適合率、再現率、F 値で評価し、生成された要約映像をアンケートによる主観評価で評価した。実験結果から、映像再生位置操作は視聴者が特に重要な注目場面を検出できると、映像再生位置操作で検出することが出来ない注目場면을停留頻度より検出することが可能であり、生成された要約映像は個人の嗜好に沿う要約映像であることを示した。

謝辞 本研究を進めるにあたり、貴重なご意見を頂いたライフスタイルデザイン研究センターの西本一志教授および宮田一乗教授に感謝の意を表します。

参考文献

- 1) 井手一郎, “映像アーカイブとファインダビリティ”, 情報の科学と技術, Vol. 60, No. 11, pp. 452-457, 2010.
- 2) T. Urruty, S. Lew, and N. Ithadaddene, "Detecting eye fixation by projection clustering", ACM Transactions Multimedia Computing, Communications and Applications, Vol. 3, No. 23, pp. 45-50, 2007.
- 3) 益満健, 越後富夫, “映像重要度を用いたパーソナライズ要約映像作成手法”, 電子情報通信学会論文誌 D, Vol. J84, No. 8, pp. 1848-1855, 2001.
- 4) 安藤昌也, 中道上, 上野秀剛, “ユーザの意図分析を可能にする映像再生位置操作記録システムの開発”, ヒューマンインタフェース学会誌, Vol. 12, No. 4, pp. 23-30, 2010.
- 5) 脇山考貴, 吉高淳夫, 平嶋宗, “注目の検出に基づいた興味モデルの作成と絵画推薦”, 情報処理学会論文誌, Vol. 48, No. 3, pp. 1048-1057, 2007.
- 6) K. Ong and W. Kameyama, "Video Summary Based on Human Bio-signals -Evaluation of its Applicability to First-Time Viewers-", IEEE Aware Computing, pp. 251-256, 2010.
- 7) H. Katti, K. Yadati, M. Kankanhalli, and C. Tat-Seng, "Affective video summarization and story board generation using Pupillary dilation and Eye gaze", IEEE International Symposium on Multimedia, pp. 319-326, 2011.
- 8) W. Peng, W. Chu, C. Chang, C. Chou, W. Huang, W. Chang, and Y. Hung, "Editing by Viewing: Automatic Home Video Summarization by Viewing Behavior Analysis", IEEE Transactions on Multimedia, Vol. 13, No. 3, pp. 539-550, 2011.
- 9) NHK Sports, “セリエ A 第 37 節 インテル 対 ナポリ”, May, 2011.