

登場人物に注目した動画内のシーン探索支援システム

檜原雅広* 松村耕平* 野間春生*

概要. 動画共有サイトが普及している。動画共有サイトの利用者は動画コンテンツを閲覧する際に、自分の興味のあるシーンのみを閲覧することや、過去に閲覧した特定のシーンを振り返ることがある。このような特定シーンを探す場面では、動画の時間的な位置を操作するシークバーを用いるが、利用者は目的の再生位置を探すために、シークバーを何度も操作しなければならないことがある。本論文では利用者がどのような情報をもとにシーン探索を行うか調査し、その結果をもとにシーン探索を支援するシステムを設計し、プロトタイプを作成した。

1 はじめに

YouTube†のような動画共有サイトが広く普及している。それらの動画共有サイトの利用率は日本、アメリカ、イギリス、ドイツ、韓国の各国で6割から7割に上る[1]。これらの動画共有サイトはテレビ放送とは異なり、利用者は自身の興味に基づいて動画閲覧をすることができる。例えば、ある動画を閲覧する際に、興味のないシーンを飛ばして、自分の好きなシーンのみを閲覧することや、過去に閲覧したことのある動画の特定のシーンを振り返ることがある。

このように、動画から特定のシーンを探すとき、利用者はシークバーを用いる。図1にYouTubeでのシークバーを用いた操作の様子を示す。シークバーは動画の時間的な位置を調整するもので、利用者はシークバーのスライダを動かすことで再生位置を調整し、サムネイルで現在の再生位置を確認することができる。しかし、シークバーを用いて特定のシーンを探すのは時間がかかる。シークバーは、動画中の時間的な位置を調整するものであるため、シーンを探すためには、利用者が探したいシーンの時間を事前に知っている必要が有るためである。すなわち、利用者はシーク操作を何度も行い、再生位置を調整することが求められる。

この問題を解決する手法としてこれまでにいくつかの研究がされてきた。Direct Manipulation Video Navigation (DMVN)は動画内の物体の動きに注目し、利用者が物体をドラッグすることで動画の再生位置を調整しようとするものである[2, 3]。例えば、動画内で車が走っているシーンで車の動く軌跡に合



図1 YouTubeでのシークバー操作

わせてポインタのドラッグ操作を行うことで再生位置を調整できる。DMVNはシーンの切り替わりがない動画については有効であるが一つの動画内でシーンの切り替わりがある場合は機能しない。一方、シーンの探索においては、切り替わりを含む動画の中から特定のシーンを探す必要がある。

本研究では、動画の中から特定のシーン探索する利用者を支援する。このために、まず、利用者がどのような情報をもとにして動画の中から特定のシーンを探索するかを調査する。そして、その調査結果をもとにして支援システムを設計・実装する。

2 調査

動画の中から特定のシーンを探すとき利用者がどのような情報を手がかりとしているのかを調査した。調査は普段から動画共有サイトを利用している大学生8人を対象に行った。

調査方法は対象者に動画の中から特定のシーンを探索するタスクを与え、その観察を行い、タスク後に半構造化インタビューを実施する。

与えたタスクは、動画の一部シーンを切り取った静止画と同じシーンを対象者に動画内から探させる

Copyright is held by the author(s).

* 立命館大学

† www.youtube.com/

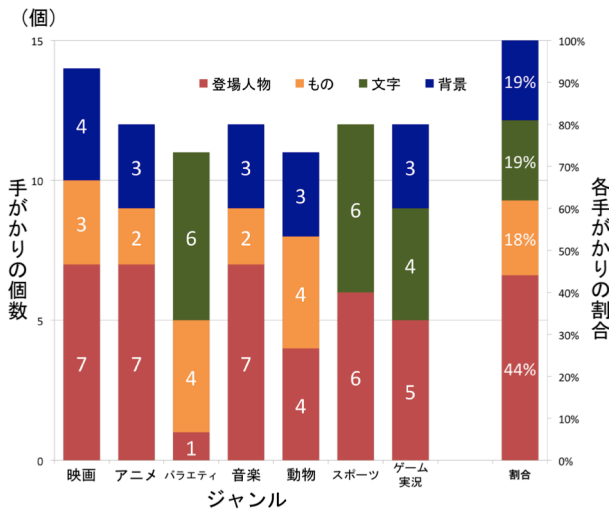


図2 ジャンルごとの手がかりの個数と割合

ことである。動画ジャンルによる違いについても検討するため、タスクでは7つの異なるジャンル（映画、アニメ、バラエティ、音楽、動物、スポーツ、ゲーム実況）の動画を準備する。タスク後のインタビューでは、「どのようにシーンを探したか」、「何を手がかりに探したか」を対象者に尋ねる。

インタビュー結果である対象者が手がかりにしたと答えたものを登場人物、もの、文字、背景の4種類に分類し、分析した。図2にジャンルごとの手がかりの個数と各手がかりの割合を表したグラフを示す。図2の左のグラフの横軸はジャンルを表し、縦軸は手がかりの個数を表している。右のグラフの縦軸は各手がかりの割合を表している。全ての手がかりのうち、44%が登場人物であったので、動画共有サイトの利用者は登場人物を主な手がかりにして、動画内のシーンを探していると判断した。

3 提案システム

調査結果から、バラエティ番組を除いた7つの動画ジャンルにおいて登場人物を手がかりにしたシーン探索が行われていることがわかった。これは、登場人物に注目して動画のシーン探索を支援することが有効であることを示唆する。

登場人物に注目したシーン探索支援を行うためには、(1) 動画中の登場人物の認識、およびその結果を反映した、(2) 動画再生インタフェースの構築が必要である。本稿では、このうち(2)のインタフェース部分について取り扱う。登場人物の認識と、登場人物と動画の紐付けを手動で行うことでプロトタイプを作成した。

提案システムでは、従来のシークバーの他に、動画を操作するインタフェースとして、登場人物ごとのタイムラインを付与する。図3にシステムのインタフェースを示す。このタイムラインは、動画内の



図3 システムのインタフェース

(動画は Tom and Jerry, 73 Episode – The Missing Mouse (1953))

各登場人物の数だけ用意され、それぞれの登場シーンにハイライトが当てられた擬似的なシークバーである。各登場人物のタイムラインは登場人物の顔と名前が書かれたボタンで表示、非表示を切り替えることができる。

4 まとめと今後の展望

本論文では、動画共有サイトの利用者がどのような情報をもとにして動画の中から特定のシーンを探索するかを調査した。さらに、その結果をもとに動画内の登場人物に注目した動画のシーン探索を支援するシステムを提案した。今後は作成したプロトタイプを用いて評価実験を行う。

また、プロトタイプでは登場人物の判別と、登場人物と動画の紐付けは手動で行っているため、システムを拡張し、自動化を図る。認識した顔画像を画像検索にかけ、検索結果をタグとして時間と紐付けることで実装する予定である。

参考文献

- [1] 総務省 | 平成 28 年版 情報通信白書 | ネット動画視聴の広がり
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nc132230.html> (2016/10 確認)
- [2] T. Karrer, M. Weiss, E. Lee, and J. Borchers, DRAGON: a direct manipulation interface for frame-accurate in-scene video navigation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08), 2008.
- [3] T. Karrer, M. Wittenhagen, and J. Borchers, DragLocks: handling temporal ambiguities in direct manipulation video navigation, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), 2012.