

伸縮タイムラインによる一人称視点映像の高速閲覧支援

樋口啓太* 米谷 竜* 佐藤 洋一*

概要. 本研究では長時間の一人称視点映像から閲覧者が高速に目的の情報へアクセスするための映像高速再生インターフェースを提案する. 一人称視点映像とは撮影者の頭部など身体に装着したウェアラブルカメラにより撮影された映像であり, 撮影者がどのような作業をしていたのかや誰と対話していたのかを詳細に記録することができる. このような映像を活用することで, 個人技能の解析や対話支援などへの応用が期待されている. しかしながら, 一人称視点映像撮影に用いるウェアラブルカメラは常時撮影が基本となるため, 映像が長時間となり閲覧したいシーンにアクセスすることに時間がかかってしまう. そこで本研究では, 閲覧者が着目する手による動作や対話など, 一人称視点映像中の手がかりを入力することにより, 解析結果から映像中の対応箇所をハイライトし, 高速再生時に自動的に速度を落とす伸縮タイムラインを搭載したインターフェースを開発する. 評価実験の結果から, 提案インターフェースにより複数または長時間の一人称視点映像からすばやく特定のイベントを発見できることを確認した.

1 はじめに

一人称視点映像は頭部や胸部などの身体に装着されたウェアラブルカメラにより撮影される映像であり, 撮影者の行動や操作, 興味対象といったものを詳細に記録することができる. このような映像を活用することで, 個人技能の解析や行動支援などへの応用が期待されている. 昨今では, スポーツや観光中のウェアラブルカメラ映像が頻繁に動画共有サービスに投稿されている. しかしながら, ウェアラブルカメラでは常時撮影が基本となるため, 記録された映像が長時間になってしまうことから, 閲覧者が必要とする情報は映像中の一部のみに記録されることとなり, 効率的にアクセスすることが難しくなってしまう. また, 一人称視点映像の特徴として, カメラが撮影者の身体に装着されているため, 映像ぶれやシーンの切り替わりが頻繁に発生するため, 従来の映像解析技術をそのまま適用することは難しい. 一人称視点映像向けの高速度再生技術 [1, 2] が研究されているが, 重要なシーンも高速に再生されるため閲覧者が看過してしまう可能性がある. そのため, 現状では閲覧者が長時間に及ぶ一人称視点映像から, どのようなユーザインターフェースによって効率的に必要な情報にアクセスできるかという問題は解決されていない.

本研究では, 長時間に及ぶ一人称視点映像から閲覧者が高速に目的となる映像を見つけることを支援する映像高速閲覧インターフェースを提案する. 本インターフェースは事前に解析した一人称視点映像中に現れる手の動作や人物との対話といった, 撮影者の行動を手がかり (一人称視点手がかり) として

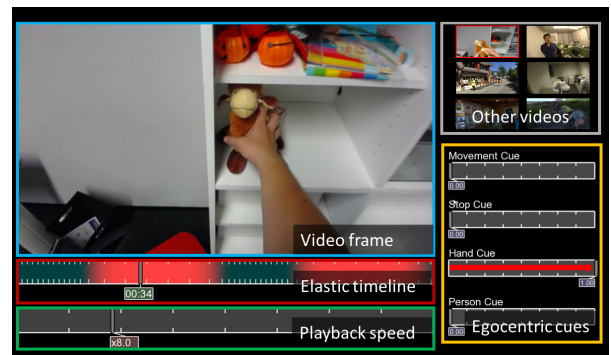


図 1. 提案インターフェース

事前に解析することにより, 重要なシーンを重点的に閲覧することを支援する. 具体的には, 閲覧者の入力した一人称視点手がかりを, 図 1 のようにビデオのタイムライン上でハイライトし伸張する. また高速再生時に, ハイライトされた箇所でのみを再生速度を落とすことにより, 高速な閲覧と重要シーンの発見を両立することができる. 図 1 の例では, 閲覧者が Hand 手がかりを入力することにより, 映像中で装着者が手を使って物体を保持している箇所がハイライトされている.

2 提案インターフェース

提案インターフェースは映像フレーム, 伸縮タイムライン, 他ビデオへのリンク, 再生速度と一人称視点手がかりを入力するスライダーから構成されている. 提案インターフェースは入力された一人称視点手がかりに応じて, 映像中の対応する箇所をハイライトする. 映像の合計再生時間は, 図 2 のように

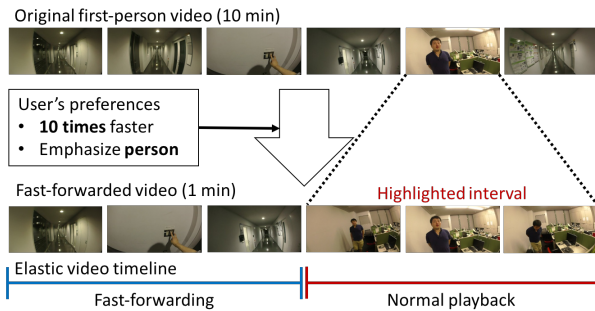


図 2. 伸縮タイムラインのコンセプト

入力された再生速度によって決まるが、ハイライトされた箇所は低速で、それ以外の箇所はより高速に再生されるように再生速度が動的に変化する。

本研究では提案インターフェースの有効性を検証するために、Stop & Movement, Hand, Person という三つの一人称視点手がかりを採用した。これらは事前に、一人称視点映像解析のためのコンピュータビジョン技術によって抽出される [3, 4, 5]。Stop & Movement 手がかりは、撮影者が静止しているか移動しているかを反映する。Hand 手がかりは映像中に撮影者の手腕が見えている箇所を、Person 手がかりは撮影者が人物との対話をしている箇所を反映することができる。閲覧者は目的となる映像に応じて一人称視点手がかりを入力することにより、高速再生時に対応するシーンを重点的に閲覧することができる。

3 評価実験

提案インターフェースの有効性を検証するために、長時間・複数の一人称視点映像からの特定イベント発見に関する評価実験を行った。本実験では、提案インターフェース (Proposed Interface) と、提案インターフェースから一人称視点手がかり入力機能を削除した Baseline Interface を比較した。実験の評価方法として、一人称視点映像群から特定のイベントを発見するタスクの完了時間を比較した。実験のための映像群として、屋内見回り (Task 1)、ナビゲーション (Task 2)、調理 (Task 3) という三つの一人称視点映像データセットを採用した。Task 1 のみ実験のために作成し、Task 2, 3 は既存のデータセットを利用した [6, 7]。被験者群はコンピュータサイエンスを専攻する大学院生・研究者 16 人から成り、実験順序が釣り合うように配置した。

図 3 の実験結果が示すように、提案インターフェースを使うことでタスク完了時間が短縮された。ウィルコクソンの符号和順位検定を適用したところ、全てのタスク及び合計時間で有意差を確認することができた。本実験の結果から、伸縮タイムラインを採

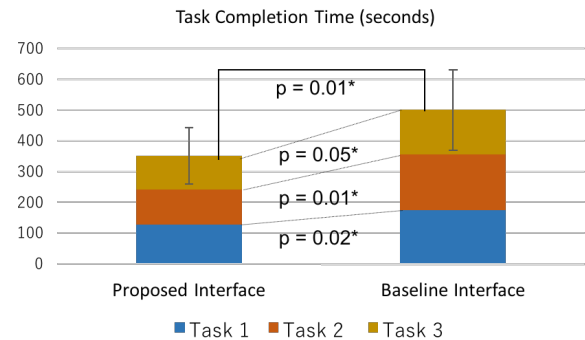


図 3. タスク完了時間の結果

用した提案インターフェースは、一人称視点映像群からの特定イベント発見を支援することができたとと言える。

4 むすび

本研究では、長時間に及ぶ一人称視点映像群から効率的に目的の情報にアクセスすることを支援する、伸縮タイムライン機能を搭載したインターフェースを提案した。本インターフェースは閲覧者が入力した一人称視点手がかりに基づき、高速再生時に映像再生速度を動的に変化させることにより、重要シーンの効率的な閲覧を実現する。評価実験により、提案インターフェースが特定イベント発見を支援することができることを明らかにした。

謝辞

本研究は JST CREST の助成を受けた。本研究の遂行にあたり助言をいただいた小山裕己氏に感謝する。

参考文献

- [1] Johannes Kopf, et al., 2014. First-person Hyper-lapse Videos, SIGGRAPH 2014.
- [2] Yair Poleg et al., EgoSampling: Fast-forward and Stereo for Egocentric Videos, CVPR 2015
- [3] Yair Poleg, et al., Temporal Segmentation of Egocentric Videos, CVPR 2014.
- [4] Cheng Li et al., Pixel-level Hand Detection in Ego-centric Videos, CVPR 2013.
- [5] Shaoqing Ren et al., Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks, NIPS 2015.
- [6] Ryo Yonetani, et al., Visual Motif Discovery via First-Person Vision, ECCV 2016.
- [7] Alireza Fathi, et al., Learning to Recognize Objects in Egocentric Activities, CVPR 2011,