7D-03

環境との相互作用を用いた一人称視点映像の検索

宮西大樹† 平山淳一郎† 前川卓也‡ 孔全‡ 守谷大樹† 須山敬之†
†国際電気通信基礎技術研究所 ‡大阪大学大学院 情報科学研究科

1 はじめに

ウェアラブルコンピューティングの進展により、日常生活中の行動だけでなく、我々が日常生活中に見ている一人称視点映像も長期に渡り計測できるようになった。この日常生活に関する一人称視点映像は長時間かつ構造化されていないため、記憶補助に利用するには、過去の一人称視点動画を検索する必要がある。従来の一人称視点動画像の検索は、手動もしくは自動でラベル付けしたテキスト [3, 4] や画像を検索クエリとして検索を行う [1]。しかし、これらの手法は記憶補助を必要とする人が忘れたことを言語で記述するか、思い出したい視覚体験と類似する画像クエリを用意しなければならない。さらに、従来の多くの手法が、日常生活動作や視覚体験と深く関係する人と環境との相互作用を考慮していない。

そこで、これらの問題を解決するため、本稿では過去の出来事を表すジェスチャを検索クエリとして用い、人と環境との相互作用を考慮しながら過去の一人称視点映像を検索する枠組みを提案する。我々の日常体験は言語化されておらず、身体の動作と密接な関係があるため、ジェスチャは過去の出来事を検索する自然な表現だと考えられる。また、ジェスチャは言語よりも容易に思い出すことができ、人種を問わず多くの人が使用できる利点がある。

2 提案手法

図 1 に提案手法の流れを掲載する。まず、本手法はユーザの日常生活に関する動作（加速度・角速度）と一人称視点映像を、ユーザが身につけたモーションセンサーとウェアラブルカメラを用いて同時に記録する。次に、人と環境との相互作用を考慮するため、記録した動作* と映像† の関係を、確率的正準相関分析（PCCA）を用いて学習する。PCCA を用いることで、両者の情報源に共通しないノイズを除去することができる。次に、過去の思い出したい出来事を表す動作をジェスチャで表現し、学習した空間中で動作と映像のペアを検索する。本検索手法は 3 秒の滑走窓を用いて 0.1 秒ずつ

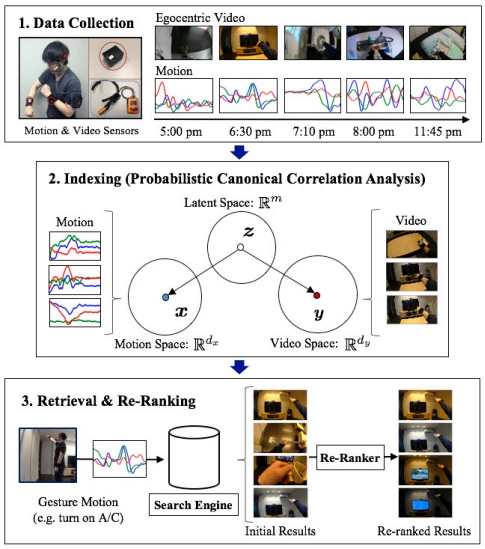

図 1: Our system retrieves videos from the egocentric video archive through a video search engine in response to a given gesture motion as a query.

ずらして全データを近傍探索するため、検索結果が冗長になる。そこで、検索した動作・映像ペアの滑走窓の時間が重複しないように検索結果の多様化を行う。さらに、検索した結果を学習した空間で疑似適合フィードバック [2] を適用することで再順位付けし、検索結果の向上を図る。

3 評価データセット

本提案システムを評価するため、被験者から実環境での日常生活動作に関する行動と映像データを収集する。本実験では、十分な量の検索対象を収集するため、実環境下で被験者が極力自然に行動する実験プロトコルを採用する。このプロトコルでは、行動と場所が記述されたワークシートを参考にして、指定された特定の場所で比較的自由に行動する。例えば、リビングルームで「TV をつける・消す」、キッチンで「冷蔵庫を開ける・閉める」などを行う。

動作データを取得するため、被験者は 3 軸加速度計と 3 軸角速度計（± 50G, ± 1500dps）を内蔵した小型かつ軽量な（40 × 30 × 20 mm, 30g）ワイヤレスモーションセンサ LP-WS1101 を頭部と両手首に装着する。また、映像データの取得するため、ウェアラブルカメラ HX-A100（1280 × 720 pixels, 29.97 fps）を

*加速度と角速度を短時間フーリエ変換を用いて周波数成分に変換し、各周波数の強度を動作特徴として用いる
†動画中の画像から深層畳込みニューラルネットによって画像特徴を抽出し、画像特徴を一定時間で重ねて動作特徴として用いる

表 1: Performance comparison of initial retrieval when using the proposed methods and baselines The best performing run is indicated in bold and statistically significant differences are marked using the symbols in the top-right corner of each method name.

Method	AP
MR$^{\heartsuit}$	0.2699
MR + PCCA	**0.3557**$^{\heartsuit}$

被験者の左側頭部に装着する。実験中は、ウェアラブルセンサを装着した被験者が各セッションで 20 種類の行動を行い、これを 10 セッション繰り返す。検索対象データの収集後、検索クエリとして使うジェスチャ動作を収集するため、前セッションで行った過去の行動を被験者に思い出してもらい、20 種類のジェスチャを 5 セッション行ってもらう。最終的に、21〜26 歳（平均：23.13、分散：1.69）の 8 人の被験者から、合計で 17 時間の行動・映像データを収集することができた。また、検索結果の適合判定のため、各行動について、ウェアラブルカメラの映像を元にして全データを人手でラベリングした。

4 実験結果

本提案システムの目的は、ジェスチャを用いた映像検索手法を用いて適合映像の一覧を返すことである。検索性能の評価には、平均精度 (AP) を用いる。平均精度は適合映像の順位での精度（検索結果中の適合映像の割合）の平均である。

提案手法の有効性を確かめるため、PCCA を行う前の空間で検索する手法 (MR) を用意した。手法間の結果を統計的に比較するため、被験者数でボンフィローニ補正した t 検定を用いる。表 1 に提案手法と比較手法の検索結果を示す。表から、PCCA で学習した空間で検索する手法 MR + PCCA を用いることで、比較手法よりも有意に平均精度を向上できることがわかった。

次に、提案システムの再順位付け手法 KDE (PCCA) の有効性を確かめる。提案システムの再順位付け手法は、PCCA で学習した空間で疑似適合フィードバックを行う。提案手法の有効性を明らかにするため、PCCA 前の空間で再順位付けする手法 KDE を比較手法として用意した。両手法とも MR + PCCA の検索結果を再順位付けした結果である。表 2 に提案手法と比較手法の検索結果を示す。表から、PCCA で学習した空間で再順位付けすることで、元の空間で検索及び再順位付けするよりも有意に平均精度を向上できることがわかった。これらの結果から、日常生活動作とその一人称視点映像ペアの検索と再順位付けには、PCCA による次元削減が有効であることがわかった。最後に、図 2 に各手法の検索結果の例を掲載する。

表 2: Performance comparison of the proposed methods and baselines when re-ranking results of MR+PCCA. The best performing run is indicated in bold and statistically significant differences are marked using the symbols in the top-right corner of each method name.

Method	AP
MR + PCCA $^{\diamond}$	0.3557
+ KDE$^{\heartsuit}$	0.3554
+ KDE (PCCA)	**0.3822**$^{\diamond\heartsuit}$

図 2: Example of search results retrieved by the methods MR, MR + PCC A, and MR + PCCA + KDE (PCCA) using gesture motions: *"clean the table"* (upper) and *"make a sandwiched cookie"* (low er). Retrieved videos are ordered by search score from left to right. Blue boxes are relevan t videos.

5 まとめ

本稿では、ジェスチャを検索クエリとする一人称視点映像検索の枠組みを提案した。提案手法は動作と映像の特徴を確率的正準相関分析を用いて獲得した空間で、ジェスチャの動作と過去の動作を比較して同時記録した映像の検索を行う。実験の結果、提案するジェスチャを用いた一人称視点映像の検索エンジンを用いることで、過去の一人称視点映像を精度良く検索できることがわかった。

参考文献

[1] V. Chandrasekhar, C. Tan, W. Min, L. Liyuan, L. Xiaoli, and L. J. Hwee. Incremental graph clustering for efficient retrieval from streaming egocentric video data. In *ICPR*, pages 2631–2636, 2014.

[2] M. Efron, J. Lin, J. He, and A. de Vries. Temporal feedback for tweet search with non-parametric density estimation. In *SIGIR*, pages 33–42, 2014.

[3] J. Gemmell, G. Bell, and R. Lueder. Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.

[4] H. Nakayama, T. Harada, and Y. Kuniyoshi. AI goggles: real-time description and retrieval in the real world with online learning. In *CRV*, pages 184–191, 2009.