

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 6 日現在

機関番号：34419

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K12028

研究課題名（和文）映像の時空間的関係を知覚可能な検索エンジンの構築

研究課題名（英文）Building a Video Search Engine based on the Perception of Spatio-temporal Relations

研究代表者

白浜 公章 (Shirahama, Kimiaki)

近畿大学・理工学部・准教授

研究者番号：30467675

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では、1. 物体の時空間的関係を考慮した映像検索、2. フレーム間の意味的な連続性に基づく映像の時間的特徴の抽出、3. 映像に対する人間の記憶伝達メカニズムを模倣するモデルの学習という3つのテーマに取り組んだ。特に、3つ目に関して、記憶を有限の外部ストレージと見立てて、映像全体の内容を最も適切に理解するために、各フレームをどれだけ記憶すべきかを判別する強化学習手法を開発した。加えて、この手法を、大量のアイテムが格納されたデータベースから、統計的に特徴的なパターンを構成するためにどのアイテムを追加・削除すべきかを判別するように拡張した手法が国際ジャーナルIEEE Accessに掲載された。

研究成果の学術的意義や社会的意義

深層学習の導入によって、画像認識性能は大幅に向上したが、映像認識では、それほどの性能向上が得られていない理由として、画像認識で用いられているたたみ込みニューラルネットワーク（CNN）の演算が人間の知覚メカニズムとよく合致している一方で、時間を伴う映像に対しては、長短期記憶（LSTM）などの既存モデルの演算が、人間の時間知覚メカニズムに合致していない点が挙げられる。この問題に対して、行動心理学に基づいて、人間の意思決定をモデル化するために有用な強化学習という手法を用いて、映像の内容を適切に理解するための記憶伝達メカニズムを模倣するモデルを学習し、その有効性を実験的に示した点に学術的意義がある。

研究成果の概要（英文）：This project addresses three main topics, 1) Video retrieval by considering spatio-temporal relations among objects, 2) Extraction of temporal features in a video by considering the continuities of semantic contents and 3) Learning a model that captures human memory mechanism for frames in a video. In particular, regarding the third topic, a reinforcement learning method has been developed to train a model that is based on a memory defined as a finite external storage and can update it so as to achieve the optimal understanding of contents in a video. In addition, this method has been extended to the data mining field, where a dataset containing a large number of items is targeted, and a model is trained to update a set of items in order to form statistically characteristic sets.

研究分野：マルチメディア情報処理

キーワード：映像検索 物体の時空間関係 グラフたたみ込み 記憶伝達 強化学習 TRECVID

1. 研究開始当初の背景

深層学習によるブレイクスルーによって、人、車、建物といった物体の認識性能は大幅に向上したが、物体の時間的、空間的關係が伴う複雑な意味を高精度に認識することは未だに困難である。特に、いかにして映像の内容の時間的關係を扱うかは重要な課題で、物体認識で大きな成果を上げているたたみ込みニューラルネットワーク (CNN) が人間の知覚メカニズムとよく合致した演算に基づいている一方で、再起型ニューラルネットワークや 3 次元たたみ込みニューラルネットワークといった、時間的關係を扱うための既存モデルの演算は、人間の時間知覚メカニズムとは大きく異なっている。この点が、深層学習を用いても、物体認識に代表される画像認識ほど、映像認識での性能向上が得られていない大きな原因だと考えられる。

2. 研究の目的

本研究では、映像の時空間的關係に対する人間の知覚メカニズムをモデル化し、特定の意味に適合する映像を高精度に検索可能なシステムを構築することを最終目的として、以下の 3 つの課題に取り組んだ。

- (1) 人間は「空間的もしくは時間的に近い物体に何らかのインタラクションを見出す」という点に着目して、物体の時空間的關係を考慮した検索手法を開発する。
- (2) フレーム間で意味が連続する (意味が類似している) 場合と不連続になる (意味が大きく変わる) 場合を区別しながら、情報 (記憶) を伝達して行き、映像の内容を表すコンパクトなベクトル表現を抽出する手法を開発する。
- (3) 上記の手法の拡張型として、人間が映像を視聴する際に無意識のうちに行っている、各フレームを記憶すべきか否かという記憶伝達メカニズムを模倣するモデルを学習する手法を開発する。

3. 研究の方法

上述の 3 つの課題に対して、以下のような具体的な手法を開発し研究を進めた。

- (1) フレームごとに物体領域を検出し、時空間的に近い位置にある領域をつないで、物体間の時空間的特徴を表すグラフを作成する。そして、グラフたたみ込みネットワークという手法を用いて、特定の意味に特有の物体間の時空間的關係を抽出し、意味を認識する手法を開発した。
- (2) 長短期記憶 (LSTM) というモデルを用いて、映像中の意味的な連続性を考慮しながら、各フレームに対する記憶を順番に伝達させて、映像全体の内容をコンパクトに圧縮したベクトルを抽出する手法を開発した。
- (3) 行動心理学に基づく強化学習という手法を用いて、記憶を有限の外部ストレージと見立てて、順番に観測されるフレームと記憶が最適にマッチするように、各フレームを重みづけして記憶に蓄積していくモデルを学習する手法を開発した。つまり、各フレームがそれまでの記憶と最適にマッチできるということは、学習したモデルによって、映像の内容を最も適切に認識できる記憶伝達メカニズムを捉えてられていることを意味する。

4. 研究成果

以下のように、3 つの課題で開発した手法の性能を評価し研究成果を得た。

- (1) 米国標準技術局 (NIST) が主催する世界的な映像解析コンテスト TRECVID 2021 の ActEV (Activities in Extended Video) 部門に参加し、5,247 本の学習用映像を用いて、36 種類のイベント (例えば、「荷物を運ぶ」、「車のドアを開ける」といった複数の物体間の特定のインタラクションによって導出される意味) を認識するモデルを学習し、3,956 本のテスト用映像を用いてイベントの認識性能を評価した。結果として、公式の評価尺度である $nAUC_{0.2}$ が 0.968 (小さいほど高性能) となり、トップの CMU が 0.423、2 位の Beijing University of Posts and Telecommunications が 0.555、3 位の University of Central Florida が 0.585 という $nAUC_{0.2}$ を達成していることと比べると、上位のチームとは大きな性能差があることが分かる結果となった。この理由として、グラフたたみ込みニューラルネットワークによって、物体ごとに周りの物体との時空間的關係を考慮した特徴が抽出できているにも関わらず、イベント認識の段階で、全ての物体の特徴を単なる平均をとって統合してしまっている点が挙げられる。

- (2) TRECVID 2008 で提供された、ニュース、ドキュメンタリー、教育番組などの様々なジャンルからの 219 本の長編映像から構成されるデータを用いて、開発した LSTM に基づく手法によってフレームごとに更新される記憶の有効性を評価した。具体的には、ベクトルとして表現された記憶と、新たに観測されたフレームの特徴を表すベクトルの類似度が高ければ、それまでの映像の内容に関する記憶と現在のフレームの内容が対応づけられた、つまり映像の内容を理解できるように適切に記憶が伝達されたということになる。結果として、開発手法による記憶の有効性は 0.778 となり、以下で述べる強化学習に基づく手法による記憶との比較に用いた。

- (3) 上記と同様の方法で、開発した強化学習に基づく手法によって更新される記憶の有効性を

評価したところ 0.798 となり、人間の行動心理学に基づく強化学習を用いた方が、より有用な記憶伝達メカニズムを捉えたモデルを学習できていることが実験的に示された。

さらに、考える全ての記憶伝達方法を検証する全探索手法、現実には不可能だが1つ先のフレームを分析した上で現在のフレームの記憶方法を決定する先読み手法とも性能比較を行い、表1の結果が得られた。まず、開発手法による記憶の伝達は、全探索手法によって求まる理論上最適な記憶伝達に匹敵することが分かった。加えて、先読み手法よりも、有効な記憶伝達が行えているということは、学習されたモデルは、1フレーム先よりもさらに先にどのようなフレームが観測されるかを予測した上で、現在のフレームをどれだけ記憶に蓄積するかを決定していることと言える。また、図1に、全探索手法、先読み手法(CSM)、開発手法による記憶伝達の例を示す。赤線で示す開発手法の記憶伝達が青線的全探索による記憶伝達と類似していることはもちろん、例えば左側の映像では、フレーム3は真っ黒である、フレーム6は既に記憶したフレーム5と類似しているため記憶する必要がない(縦軸の値がゼロ)、右側の映像では、フレーム2は今後のシーンの状況を表す、フレーム5は初めて人が出現したため記憶する必要がある(縦軸の値が大)というように、意味的にも妥当な記憶伝達が行えていることが分かる。

表 1 : 異なる手法によって伝達された記憶の有効性の比較

	全探索手法	先読み手法	強化学習
記憶の有効性	0.6322	0.6254	0.6302

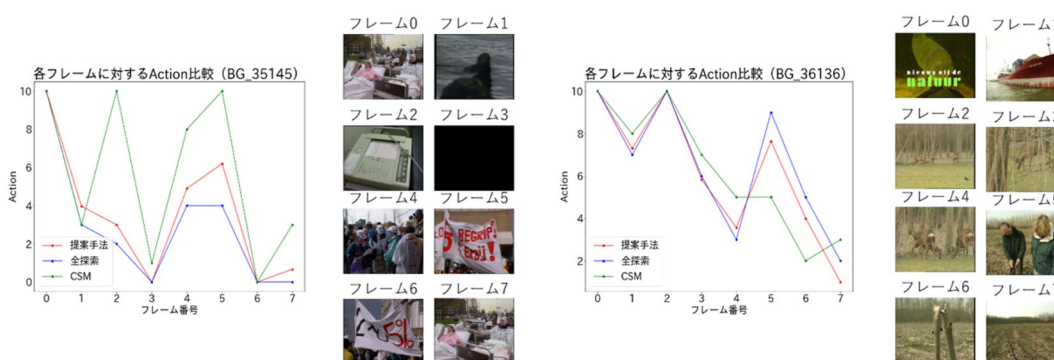


図 1 : 異なる手法による記憶伝達の例

加えて、学習用映像を増やすと、より有用な記憶伝達メカニズムを捉えたモデルを学習できることが実験的に示されており、本研究で使用したよりもさらに大規模な学習用映像を用いれば、様々な映像に対して汎用的に適切な記憶伝達を行えるモデルを構築できることが示唆された。最後に、本手法を2021年10月の国内研究会で発表したところ、幹事の方から、大きな技術的新規性が認められるとのコメントをいただいた。

(4) 上記で開発した強化学習に基づく記憶伝達手法を、データマイニングの分野に応用・拡張した。概要として、大量のアイテムが格納されたデータベースから、統計的に特徴的なアイテムの集合を抽出するアイテム集合マイニングを対象として、抽出したいアイテム集合の種類を指定してやれば、現在吟味しているアイテム集合に対して適切なアイテムを追加・削除して、指定した種類のアイテム集合を形成するモデルを学習する手法を開発した。つまり、映像では、将来どのようなフレームが観測されるかを予測した上で現在のフレームをどれだけ記憶に蓄積すべきかを推定するモデルを学習した手法を、将来的に指定された種類のアイテム集合を形成できるかどうかを予測した上で、現在吟味しているアイテム集合にどのアイテムを追加もしくは削除すべきかを推定するモデルを学習できるように拡張したということである。実験では、高効用アイテム集合、頻出アイテム集合、相関ルールという3種類のアイテム集合を対象として、抽出すべきアイテム集合のうちの83.2%から100%を抽出できるモデルを学習することに成功した。ここで、既存のアイテム集合マイニング手法では、抽出するアイテム集合の種類ごとに、アルゴリズムやデータ構造を実装し直さなければならないという汎用性のなさが問題となっていた。これに対して、開発手法には、アイテム集合の種類(正確には、アイテム集合の評価値の算出方法)さえ指定してやれば、その種類のアイテム集合のほとんどを抽出可能なモデルを自動的に学習できるという重要な学術的意義がある。この研究成果は、世界的な国際ジャーナル IEEE Access に掲載されている。

加えて、映像という時系列データの時間的関係を考慮すること得た知見に基づいて、ウェアラブル端末に搭載されたセンサ(加速度計や角速度計など)からの時系列データから、人間の日常行動を認識する手法を開発し、その成果が Sensors という国際ジャーナルに掲載されている。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 2件/うちオープンアクセス 3件）

1. 著者名 Nisar Muhammad Adeel, Shirahama Kimiaki, Li Fr?d?ric, Huang Xinyu, Grzegorzec Marcin	4. 巻 20 (Article No. 3463)
2. 論文標題 Rank Pooling Approach for Wearable Sensor-Based ADLs Recognition	5. 発行年 2020年
3. 雑誌名 Sensors	6. 最初と最後の頁 1~21
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/s20123463	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Frederic Li, Shirahama Kimiaki, Nisar Muhammad Adeel, Huang Xinyu, Grzegorzec Marcin	4. 巻 20 (Article No. 4271)
2. 論文標題 Deep Transfer Learning for Time Series Data Based on Sensor Modality Classification	5. 発行年 2020年
3. 雑誌名 Sensors	6. 最初と最後の頁 1~25
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/s20154271	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Fujioka Kazuma, Shirahama Kimiaki	4. 巻 10
2. 論文標題 Generic Itemset Mining Based on Reinforcement Learning	5. 発行年 2022年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 5824~5841
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2022.3141806	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件/うち国際学会 3件）

1. 発表者名 Daiki Mukai, Ryosuke Utsunomiya, Shunsuke Utsuki, Kimiaki Shirahama, Takashi Matsubara and Kuniaki Uehara
2. 発表標題 Kindai University and Osaka Gakuin University at TRECVID 2020 AVS and ActEV Tasks
3. 学会等名 TREC Video Retrieval Evaluation (TRECVID) 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 西本光希, 佐々木海斗, 白浜公章
2. 発表標題 個人差を考慮したスライドに対する瞬き率推定
3. 学会等名 2021年電子情報通信学会総合大会
4. 発表年 2021年

1. 発表者名 向井大貴, 野見山哲平, 白浜公章
2. 発表標題 マルチタスク学習を用いたセンサデータからの行動認識
3. 学会等名 2021年電子情報通信学会総合大会 (国際学会)
4. 発表年 2021年

1. 発表者名 Naoki Takashima, Frederic Li, Marcin Grzegorzec and Kimiaki Shirahama
2. 発表標題 Cross-modal Music-emotion Retrieval Using DeepCCA
3. 学会等名 The Eighth International Conference on Information Technology in Biomedicine (ITIB 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 西本光希, 白浜公章
2. 発表標題 強化学習を用いた映像フレームに対する記憶メカニズムの獲得
3. 学会等名 映像表現&コンピュータグラフィックス研究会 (AIT)
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Deep Transfer Learning for Time Series Data
<https://www.info.kindai.ac.jp/~shirahama/transfer/>
Sensorsで発表した手法のソースコード，及び実験データを公開している．

Rank Pooling Approach for ADLs Recognition
https://www.info.kindai.ac.jp/~shirahama/rank_pooling/
Sensorsで発表した手法のソースコード，及び実験データを公開している．

Wisteria30 / GIM-RL
<https://github.com/Wisteria30/GIM-RL>
IEEE Accessで発表した強化学習手法のソースコード，及び実験データのダウンロード方法や使用方法を公開している．

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
ドイツ	リューベック大学			