

令和 6 年 6 月 21 日現在

機関番号：82626

研究種目：基盤研究(A)（一般）

研究期間：2021～2023

課題番号：21H04908

研究課題名（和文）場の文脈を考慮した異常行動の自動検知

研究課題名（英文）Out-of-context Action Detection

研究代表者

原 健翔（Hara, Kensho）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：70837575

交付決定額（研究期間全体）：（直接経費） 25,300,000円

研究成果の概要（和文）：本研究では、動画像における文脈をモデル化し、文脈を考慮した異常検知技術を確立することを目的とし、(a)空間的な文脈のモデル化及び(b)時間的な文脈のモデル化、(c)文脈にとらわれることなく人物行動を表現するモデル、(d)文脈を考慮した異常行動の自動検出アルゴリズムの研究に取り組んだ。昨今の大規模言語モデルの発展により(d)については一定のレベルでの解決が見えてきたこともあり、特に(a)-(c)に注力して研究を進めた結果、シーンの空間的・時間的な文脈を表現するためのシーングラフ構造の検出手法の確立を始めとして、文脈及び人物行動を詳細に表現するためのモデルの構築を進めることができた。

研究成果の学術的意義や社会的意義

近年の研究開発により人工知能技術が急速に進展により視覚的な情報に基づく単純な識別問題は解決されつつある一方で、動画像中の場面や状況を表現する「文脈」を考慮し動画像をより高度に解析することは困難な問題として残っている。本研究では人物と周囲の物体や環境との関係性を捉えることが文脈の表現に重要であることに着目し動画中の文脈の表現を可能とした。これにより、大規模言語モデルの文脈理解能力と組み合わせることで動画中の文脈に従った人物行動の異常検知を可能とし、より高度な動画像解析の実現及び人々の安全・安心を支える基盤技術の構築に貢献した点に学術的・社会的意義がある。

研究成果の概要（英文）：The purpose of this research is to model the context in videos and establish an anomaly detection framework based on the context. Specifically, we focused on (a) modeling spatial context, (b) modeling temporal context, (c) developing models that represent human behavior independently of context, and (d) researching automatic detection algorithms for anomalous behavior based on the context. Since the problem of (d) is being resolved to some extent given the recent advancements in large-scale language models, we concentrated our research efforts particularly on (a) through (c). As a result, we have made progress in constructing models that represent context and human behavior in details, including the establishment of a scene graph detection method to represent the spatial and temporal context of scenes.

研究分野：コンピュータビジョン

キーワード：深層学習 行動認識 動画認識 Transformer シーングラフ

1. 研究開始当初の背景

COVID-19 の流行もあり社会的に無人店舗の重要性増加など、防犯・防災の監視に穴が空きやすいうちで、人々の安全・安心を支えていくための技術として AI による異常行動の自動検知の重要性が増していた。異常検知を始めとする動画画像解析において重要となるのが、動画画像中の場面や状況を表現する「文脈」である。実世界において生じる事象は、その時点の文脈に強く依存するためである。異常検知においても、ある行動が異常かは、その文脈に依存して決まる。例えば「施錠せずに外出する」という行動を考えた時に外に出たのが自宅からであるから施錠していないことが異常であり他人の家や店から外に出るときに何もしないのは正常である。しかし、動画画像のような時系列的に変化する対象においては文脈もそれに従い変化するため定義が難しいこともあり、従来の動画画像解析では文脈を明示的に考慮することなく、動画中の事象を解析することに注力していた。従って、動画画像解析の学術的な発展のためには、文脈を定義してモデル化し、事象と文脈を明示的に切り分けて解析する技術の確立が期待される状況にあった。

2. 研究の目的

上述した背景を踏まえて、本研究では動画画像における文脈をモデル化し、文脈を考慮した異常検知技術の確立を目的として設定した。文脈については、動画画像が時系列的に変化することを踏まえ、空間的な文脈（ある時点でどういう状況か）と時間的な文脈（過去からの変化を踏まえてどのような状況か）という二つの側面からモデル化することを試みた。近年の深層学習の発展により、動画画像中の人物行動の認識といった、事象の解析技術は成熟しつつある一方で、事象の背景の文脈の解析は見落とされている技術であった。本研究ではそこに焦点を当て、動画画像中の事象と文脈の両方を解析可能とすることで、本研究で取り組む異常検知のみならず、文脈を考慮した行動意図の推定など、より高度な動画画像解析の実現に貢献することを目指した。

3. 研究の方法

本研究では、(a)空間的な文脈のモデル化及び(b)時間的な文脈のモデル化により、動画画像中に現れる文脈を表現するモデルを構築した。また、(c)文脈にとらわれることなく人物行動を表現するモデルについても構築した。これらを基に、(d)文脈を考慮した異常行動の自動検出アルゴリズムを確立することを目指した。しかしながら、研究を開始してから ChatGPT を始めとした大規模言語モデルの急速な進展があり、文脈と人物行動を自然言語で記述し入力することで、その行動が正常か異常かを判断する点については現実的なものとなりつつある状況になった。そこで、特に(a)～(c)に注力し、動画中の文脈のモデル化及び人物行動の表現の獲得を進めていった。各項目の詳細を以下に述べる。

(a) 空間的な文脈のモデル化

空間的な文脈として、本研究では人物と物体の関係性に着目した。人物の行動と周囲の物体は密接に関係しているため、人物行動に着目した上である時点でどういう状況かを表現するためにはある人物が周囲の物体のどのような関係にあるのかを理解することが重要となる。そこで、本研究では動画中の人物と物体の関係性を表現するシーングラフを生成する手法について研究した。従来手法では物体を検出した後に関係性を検出しており、冗長な推論であり処理のボトルネックになっていることに加えて、関係性検出の性能が物体検出の性能に依存するという問題点があった。本研究では、物体と関係性を同時に検出するための手法を提案することでこの問題の解決を図った。物体と関係性の相互関係を学習する Graph Tensor Decoder というモジュールを Transformer に導入することで、物体・関係性の同時検出を実現した。これにより、従来手法に比べて約 3 倍の推論速度の実現と同時にメモリ消費量を約 1/3 に削減しながら、信頼度の高い人物・物体関係性検出を実現した。本成果は SSII2022 で最優秀学術賞を受賞するなど高く評価された。

(b) 時間的な文脈のモデル化

時間的な文脈のモデル化の 1 つとして、(a) で研究した人物と物体の関係性の表現に、時間的な文脈を組み込む手法について研究した。動画中の過去の系列で人物と物体の関係性がどのような状況にあったのかという情報は現在の関係性に影響するものであるため、過去の関係性を考慮して現在の関係性を検出することが重要となる。そこで、(a) で提案した Graph Tensor Decoder が出力する Graph Tensor の過去の系列を集約する Temporal Context Aggregator を導入する手法を提案した。これにより、より高い性能で人物・物体の関係性を検出することを実現した。図 1 に関係性検出の例を示す。本成果は、DIA2023 で研究奨励賞を受賞するなど高く評価された。

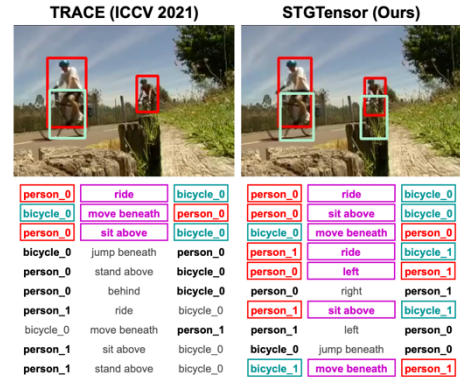


図 1 人物・物体間の関係性検出の例

また、過去の時系列に基づき未来の人物行動を予測する手法についても研究した。現在行なっている行動から未来の行動へ遷移する際の振る舞いを認識することで未来の行動を予測することができるという発想に基づき、行動間の遷移をカテゴリとして定義し認識するモデルを学習することで、未来の行動を安定して予測する手法を提案した。

加えて、シーン中の物体の状態の変化を高度に理解するために、画像中に含まれる物体が自然に風化する様子を合成する手法を研究した。風化を表現するテクスチャを物体に対して自然に合成する手法を提案し、風化が進んでいく様子を動画として生成することを可能とした。また、物体状態を自由に变化させるためのテクスチャを生成する手法についても研究し、VC2023 にて企業賞を受賞及した他トップ国際会議である SIGGRAPH Asia 2023 に採択された。

(c) 動きのみに注目した人物行動モデル

動きのみに注目して人物行動をモデル化する 1 つのアプローチとして人物の姿勢に着目したものがあ。人物姿勢の遷移に基づいて人物行動を表現することで、周囲の物体や環境とは独立して人物行動を表すことが可能となる。人物姿勢に基づいて人物行動を正確に表現するために、人物姿勢の推定手法の改善に取り組んだ。人口統計学的バイアスを考慮することで多様な体型の人物に対して高精度に体型と姿勢の推定を可能とした。本研究は MIRU2022 でインタラクティブ発表賞を受賞するなど高く評価された。

(d) 文脈を考慮した異常行動の検出

文脈を考慮した異常行動の検出に対するアプローチとして、1 つは近年急速に進展している大規模言語モデルを活用する方法が挙げられる。文脈と行動をテキストとして記述し大規模言語モデルにプロンプトとして入力し、その行動が正常か異常かを回答させることで異常行動を検出することができる。実際に ChatGPT を用いてこのようなアプローチを試したところ、一般常識に基づいて判断できるような異常については適切に検出することが確認できた。

また、別のアプローチとして、人物の骨格という行動の情報に加えて人物に付随する物体という文脈の情報を利用した異常検知の手法についても研究した。骨格情報の推定と物体検出を同時に行い、両者の情報を統合した認識モデルにより異常行動を検知することを可能とした。

4. 研究成果

本研究を通して、シーンの空間的・時間的な文脈を表現するためのシーングラフ構造の検出手法の確立を始めとして、文脈及び人物行動を詳細に表現するためのモデルの構築を進めることができた。これらにより動画中の人物行動および文脈を認識することで自然言語の形式で動画を記述することが可能となる。これらの記述に基づいて大規模言語モデルを活用することで、文脈に基づいた異常行動検出の実現が期待される。また、大規模言語モデルを活用することを前提とした場合には、動画中の情報をユーザの要求に従って適切な粒度で記述する技術が重要となると考えられる。そのような技術については 2024 年度より採択された科研費基盤 B「動画像を目的に応じた情報粒度で言語記述する AI の構築」を進めていくことを予定している。

論文などの研究成果としては、トップ国際会議の 1 つである SIGGRAPH ASIA 2023 への採択に加えて、第 25 回画像の認識・理解シンポジウム (MIRU2022) でインタラクティブ発表賞を受賞、動的画像処理実利用化ワークショップ 2023 (DIA2023) で 2 件研究奨励賞を受賞、2023 年度精密

工学会研究奨励賞を受賞，第 28 回画像センシングシンポジウム (SSII2022) 最優秀学術賞を受賞など，多数の受賞により実施した研究が高く評価された。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 川野 恭史, 長崎 好輝, 原 健翔, 青木 義満, 片岡 裕雄	4. 巻 89
2. 論文標題 予備動作の認識による行動予測	5. 発行年 2023年
3. 雑誌名 精密工学会誌	6. 最初と最後の頁 915 ~ 920
掲載論文のDOI (デジタルオブジェクト識別子) 10.2493/jjspe.89.915	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 -

1. 著者名 森山 凌, 金子 直史, 鷺見 和彦	4. 巻 89
2. 論文標題 監視カメラ映像における物体情報を付与した骨格ベースの異常行動検知	5. 発行年 2023年
3. 雑誌名 精密工学会誌	6. 最初と最後の頁 934 ~ 941
掲載論文のDOI (デジタルオブジェクト識別子) 10.2493/jjspe.89.934	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 -

1. 著者名 Hao Guoqing, Iizuka Satoshi, Hara Kensho, Kataoka Hirokatsu, Fukui Kazuhiro	4. 巻 11
2. 論文標題 Natural Image Decay With a Decay Effects Generator	5. 発行年 2023年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 120402 ~ 120418
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2023.3328171	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計13件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Yasufumi Kawano, Yoshiki Nagasaki, Kensho Hara, Yoshimitsu Aoki, Hirokatsu Kataoka
2. 発表標題 Unsupervised Semantic Segmentation Leveraging Video Consistency
3. 学会等名 第26回画像の認識・理解シンポジウム MIRU2023
4. 発表年 2023年

1. 発表者名 森山 凌, 金子 直史, 鷺見 和彦
2. 発表標題 場の文脈を考慮した異常検知に向けたコンテキスト生成手法の検証
3. 学会等名 精密工学会 画像応用技術専門委員会 サマーセミナー-2023
4. 発表年 2023年

1. 発表者名 Guoqing Hao, Satoshi Iizuka, Kensho Hara, Edgar Simo-Serra, Hirokatsu Kataoka, Kazuhiro Fukui
2. 発表標題 Diffusion-based Holistic Texture Rectification and Synthesis
3. 学会等名 Visual Computing 2023
4. 発表年 2023年

1. 発表者名 Guoqing Hao, Satoshi Iizuka, Kensho Hara, Edgar Simo-Serra, Hirokatsu Kataoka, Kazuhiro Fukui
2. 発表標題 Diffusion-based Holistic Texture Rectification and Synthesis
3. 学会等名 SIGGRAPH Asia 2023 (国際学会)
4. 発表年 2023年

1. 発表者名 金子 直史, 丸山 楽斗, 伊東 聖矢, 鷺見 和彦
2. 発表標題 人口統計学的バイアスを考慮した人物の3次元姿勢体型推定
3. 学会等名 第25回 画像の認識・理解シンポジウム (MIRU2022)
4. 発表年 2022年

1. 発表者名 Guoqing Hao, Satoshi Iizuka, Kensho Hara, Hirokatsu Kataoka, Kazuhiro Fukui
2. 発表標題 Natural Image Weathering with a Generative Decay Network
3. 学会等名 Visual Computing 2022
4. 発表年 2022年

1. 発表者名 川野 恭史, 長崎 好輝, 原 健翔, 青木 義満, 片岡 裕雄
2. 発表標題 予備動作の認識による行動予測
3. 学会等名 ビジョン技術の実利用ワークショップ ViEW 2022
4. 発表年 2022年

1. 発表者名 長崎 好輝, 川野 恭史, 山本 晋太郎, Qiu Yue, 青木 義満, 原 健翔, 片岡 裕雄
2. 発表標題 動的シーングラフ生成におけるGraph Tensorを用いた時空間情報の抽出
3. 学会等名 ビジョン技術の実利用ワークショップ ViEW 2022
4. 発表年 2022年

1. 発表者名 長崎好輝, 川野恭史, 山本晋太郎, Yue Qiu, 青木義満, 原健翔, 片岡雄裕
2. 発表標題 動的シーングラフ生成における網羅的かつ文脈レベルでの時空間情報抽出
3. 学会等名 動的画像処理実利用化ワークショップ DIA 2023
4. 発表年 2023年

1. 発表者名 森山 凌, 金子 直史, 鷲見 和彦
2. 発表標題 監視カメラ映像における物体情報を付与した骨格ベースの異常行動検知
3. 学会等名 動的画像処理実利用化ワークショップ DIA 2023
4. 発表年 2023年

1. 発表者名 長崎 好輝, Qiu Yue, 青木 義満, 原 健翔, 片岡 裕雄
2. 発表標題 シーングラフ生成におけるTransformerを用いたEnd-to-Endモデルの提案
3. 学会等名 ビジョン技術の実利用ワークショップ ViEW 2021
4. 発表年 2021年

1. 発表者名 川野 恭史, 原 健翔, 青木 義満, 片岡 裕雄
2. 発表標題 行動遷移に着目した行動予測
3. 学会等名 ビジョン技術の実利用ワークショップ ViEW 2021
4. 発表年 2021年

1. 発表者名 長崎 好輝, 川野 恭史, 山本 晋太郎, Qiu Yue, 原 健翔, 青木 義満, 片岡 裕雄
2. 発表標題 動的シーングラフ生成における物体と関係性の同時検出
3. 学会等名 第28回画像センシングシンポジウム SSII 2022
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

DIA2023研究奨励賞, 長崎好輝, 川野恭史, 山本晋太郎, Yue Qiu, 青木義満, 原健翔, 片岡裕雄「動的シーングラフ生成における網羅的かつ文脈レベルでの時空間情報抽出」
DIA2023研究奨励賞, 森山凌, 金子直史, 鷲見和彦, 「監視カメラ映像における物体情報を付与した骨格ベースの異常行動検知」
SSI12022 最優秀学術賞, 長崎好輝, 川野恭史, 山本晋太郎, Qiu Yue, 原 健翔, 青木義満, 片岡裕雄, 「動的シーングラフ生成における物体と関係性の同時検出」
VC2023 企業賞, Guoqing Hao, Satoshi Iizuka, Kensho Hara, Edgar Simo-Serra, Hirokatsu Kataoka, Kazuhiro Fukui, "Diffusion-based Holistic Texture Rectification and Synthesis"

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
研究分担者	青木 義満 (Aoki Yoshimitsu) (00318792)	慶應義塾大学・理工学部(矢上)・教授 (32612)	
研究分担者	金子 直史 (Kaneko Naoshi) (40803531)	東京電機大学・未来科学部・准教授 (32657)	
研究分担者	片岡 裕雄 (Kataoka Hirokatsu) (70784883)	国立研究開発法人産業技術総合研究所・情報・人間工学領域・上級主任研究員 (82626)	

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
---------	---------