

令和元年5月29日現在

機関番号：13901

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K12714

研究課題名(和文) 実用的なビデオ暴力度のレーティング技術に関する研究

研究課題名(英文) A Study on Practical Visual Violence Rating

研究代表者

ワン ユ(WANG, YU)

名古屋大学・国際開発研究科・助教

研究者番号：60724169

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究は、実用性の高い、高性能のビデオ暴力度レーティング手法を開発するため、(1)人物検出に基づいた時空間特徴プーリング技術、(2)グラフィック・モデルを用いた暴力要素認識結果の最適化技術、(3)リストワイズ・ランキングに基づいた暴力度レーティング技術及び、(4)時空間並列処理技術の研究開発を行いました。また、これらの要素技術を評価するために、大規模なビデオデータを収集・ラベリングし、定量的な評価を行いました。先行研究より高いレーティング精度及び処理速度が確認でき、暴力度レーティング技術の実用化に向けて、一歩近づくことが出来ました。

研究成果の学術的意義や社会的意義

本研究で開発した暴力度レーティング技術は実用性が高く、実用化出来れば、従来ほとんど手動で行われてきた膨大な数のビデオコンテンツの分類を自動化できる。また、この技術により推定した暴力度から、ビデオごとに視聴に適した年齢層を見積もることができるため、青少年をはじめとする多くの利用者にとって、より自由で安全な視聴空間が確保できる。本研究で暴力度レーティングの実用化のために開発した要素技術は、暴力度レーティングの精度と速度を大幅に向上させることが出来た他、映像認識タスクにおいて最も基本的な課題の解決にも貢献できる。研究開発成果はビデオ分類、要約、検索等の研究にも大きな推進力を与えることになる。

研究成果の概要(英文)：In this work, in order to develop high performance practical visual violence rating approach, we worked on the following four aspects: (1) human detection-based spatial-temporal feature pooling methods; (2) graphical model-based methods for optimizing violent attribute estimations; (3) list-wise ranking based violence rating methods; and (4) spatial-temporal parallel processing methods. In order to evaluate these methods, we collected and labeled a large number of video clips, and conducted experiments quantitatively. We confirmed that these methods lead to significant improvements on rating speed and rating accuracy over the previous works, and push the visual violence rating technique one step forward toward the practical usage.

研究分野：画像処理

キーワード：暴力度レーティング 視覚的暴力 行動認識

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

ビデオコンテンツに含まれている暴力・残虐表現が、成長過程の青少年にネガティブな影響を与えることは、これまで社会学者、社会心理学者によって実証されてきた。暴力・残虐表現を含む有害コンテンツから青少年を守るために、放送事業者、視聴者団体、行政の三者は連携して、放送時間帯の配慮、番組情報の公開、保護者の番組選択を支援するVチップ技術の提供など、様々な措置を取ってきた。しかし、近年、インターネットにつながるスマートテレビや電子機器の浸透等により、青少年がウェブ上に散在するビデオコンテンツにアクセスする手段は、多様化・パーソナル化しており、視聴内容のコントロールは困難になっている。そのため、従来の三者連携による施策は、実質的に効力を失いつつある。有害コンテンツの自動解析・フィルタリング技術が強く求められている。

動画像解析分野では、映像特徴を利用した暴力コンテンツの自動解析・フィルタリングは、まだ本格的な実用段階に達していないが、いくつかのアプローチが試みられた。既存研究は、いずれも計算速度が不十分な上、(1)ビデオに含まれる単独の暴力シーンの検出を目標としており、ビデオ全体に対する評価を行っていない；(2)映像フレームにおける暴力シーンが存在するか否かの二者択一の分類問題にとどまり、「暴力度(レベル)」の評価が得られないなど、実用的な完成度にはまた達していない。研究代表者は、実用的な暴力コンテンツの自動フィルタリング技術を確立することを目指して、これまで「ビデオ単位の暴力度レーティング」というコンセプトを提案し、初となる暴力度自動レーティング手法(先行研究)を開発した。既存研究と比べて、先行研究は、(1)シーンレベルの解析ではなく、ビデオレベルの解析である；(2)暴力シーンの検出ではなく、暴力レベルの強弱を示す暴力度の推定である。先行研究の出力となる暴力度から、入力ビデオの視聴に適した年齢層を直接に見積もることが可能であるため、従来ほとんど手動で行ってきたビデオコンテンツの分類を、効率的に行うことができる。また、近い将来の実用を見据えて、利用形態についても研究調査を実施した。しかし、現状では、先行研究の精度はまだ低く、処理速度も実時間より大幅に遅れるといった課題がある。実用化のため、これらの課題を早急に解決しなければならない。本研究では、先行研究の提案手法の精度と処理速度を、実用に十分に耐えられる水準まで向上させるための新たな技術を研究開発する。

2. 研究の目的

本研究は、実用性の高い、高性能のビデオ暴力度レーティング手法を開発することを目的とする。個々の入力ビデオに対し、その内容に基づき暴力度を推定して出力し、出力した暴力度は、ビデオごとに視聴に適した年齢層の見積もりができるよう定義されている。インターネット上で急増しているワイルドなビデオコンテンツに対応可能であるため、次世代テレビ視聴やスマートフォン・タブレットでの視聴にも適合した、より安心・安全なインターネット環境の実現に貢献することができる。本研究では、研究代表者が行った先行研究に、新しい特徴抽出技術、機械学習技術及び並列処理技術を導入することにより、暴力度レーティングの精度と速度を、実用レベルまたは実用レベルに近い水準まで向上させる。具体的には、(1)人物検出に基づいた時空間特徴プーリング技術、(2)グラフィック・モデルを用いた暴力要素認識結果の最適化技術、(3)リストワイズ・ランキングに基づいた暴力度レーティング手法の開発、及び、(4)時空間並列処理の導入という4つの研究項目により実現する。

3. 研究の方法

以下、本研究の研究方法について詳述する。

[1] 人物検出に基づいた時空間特徴プーリング技術の開発

ウェブ上に散在するビデオコンテンツは様々な条件・環境の下で撮影されており、背景雑音やカメラ移動が多いため、特徴抽出の精度が大きく下がる。そこで、本研究では、すべての映像フレームにおいて高速人物検出器を動作させ、各ショットに対応する時空間を、人物が存在する時空間断片と存在しない時空間断片に分類し、それぞれの時空間断片で特徴量のプーリング(特徴の集計方法)を行い、特徴ベクトルを得る。このようにして、人物動作に属する特徴と背景雑音やカメラ移動に属する特徴を区別することで、暴力要素の認識精度および暴力度レーティングの精度を大幅に向上させることを狙う。

[2] グラフィック・モデルによる暴力要素認識結果の最適化技術の開発

暴力要素の認識は、映像中の武器、爆発、殴打などのシーンの認識を指す。この認識性能は、あとの暴力度レーティングの精度を大きく左右する。本研究では、映像特徴から暴力要素を認識する学習器の開発と同時に、暴力要素間の確率的依存関係も学習する。すなわち、学習データから、暴力要素間の条件確率(構造)を、グラフィック・モデルとして学習し表現する。新たなビデオが入力されたとき、まず映像特徴のみを使って暴力要素を認識する。認識結果となる暴力要素それぞれの存在確率は、さらに要素間の依存関係を表現するグラフィック・モデルの下で更新する。この処理は、画像特徴よりハイレベルの知識に基づいた最適化であるため、暴力要素の認識精度が大幅に向上されることが期待できる。

[3] リストワイズ・ランキングに基づいた暴力度レーティング手法の開発

これまで開発したペアワイズ・ランキングに基づいた暴力度レーティング手法は、任意の2つのデータポイントを比較するためのモデルを学習するものである。未知のデータポイントのランクを予測する時、既存のデータポイントとの1対1の強弱関係が推定されることで、未知のデータポイントのランクが推定される。しかし、ペアワイズ・ランキングは学習データのローカルな情報に基づいているため、既知のデータポイントの中に、未知のデータポイントとのランク差が大きいものが多く存在する場合、見積もりの信頼性は低い。

一方、リストワイズ・ランキング手法は、任意のデータポイントのランキング・スコアを計算するランキング関数（モデル）を学習する。未知のデータポイントのランクを予測する時、ランキング関数を用いてランキング・スコアが得られる。しかし、リストワイズ・ランキングはデータのグローバルな情報に基づいているため、学習段階でのモデルの最適化が困難であり、精度の高いモデルを学習することが難しい。申請者は、両手法の特徴を相互に補い合うように統合し、粗密（coarse-to-fine）スタイルの暴力度レーティング手法を開発する。すなわち、まず、リストワイズ・ランキング手法で未知のビデオコンテンツのランキング・スコア（大まかな位置）を算出し、次に、このスコアの周辺値をもつ既知のビデオコンテンツと、ペアワイズ・ランキング手法により比較することで、未知のビデオコンテンツの暴力度をより正しく確実に予測する。

[4] 時空間並列映像処理の導入

本研究において処理コストが最も大きいのは、ショット分割のためのフレーム上の空間特徴計算及び暴力要素認識時の時空間特徴の抽出である。前者は連続フレーム間の色やエッジなどの変化を捉える特徴の計算であるため、時間上の分散処理を提案する。一方、後者は時空間ボリュームにおける局所時空間特徴の計算である。前者と同じように、各時空間ボリュームの特徴計算は並列処理できる独立なGPUのコアに実行させる。

4. 研究成果

研究項目[1]については、当初提案していた人物検出に基づいた時空間特徴プーリングというシャローラーニング手法を開発しましたが、高い有効性は確認できませんでした。そこで、研究方針を変更し、同じアイデアをベースにSTN-CNNというデープランニング手法を開発し、人物領域の時空間特徴プーリングを実現しました。詳細行動認識タスクにおいて、高い有効性を確認できました。また、STN-CNNを一括に学習するのは困難であるため、STNの学習のための教師ありプレトレーニング手法を開発し、学習過程を安定化させた。

研究項目[2]については、当初提案していたグラフィック・モデルを用いた認識結果の最適化技術の開発を行いました。また、提案モデルの有効性を確認するために、新たに1,930個のビデオデータを収集・ラベリングし、評価実験を行いました。

研究項目[3]については、リストワイズ・ランキングに基づいた暴力度レーティング手法を開発し、ある程度有効性を確認できた。レーティング精度をさらに向上させるため、予定していた計画と別に、Two-stream CNNとペアワイズ・ランキングを組み合わせた新しい手法を提案し、大幅の性能向上を確認出来た。

研究項目[4]については、IDTなどの時空間特徴をCNN時空間特徴に変更し、複数枚GPUを同時利用した特徴抽出手法を実装し、高い特徴抽出速度を確認できました。

5. 主な発表論文等

〔雑誌論文〕(計1件)

[1] Guanwen Zhang, Jien Kato, Yu Wang, Kenji Mase, "A Novel Approach for Annotation-based Image Retrieval Using Deep Architecture", Journal of Multiple-Valued Logic and Soft Computing, Vol.30, No. 4-6, PP. 541-558, 2018. (査読有)

〔学会発表〕(計13件)

[1] Ying Ji, Yu Wang and Jien Kato, "Visual Violence Rating with Pairwise Comparison", The 26th IEEE International Conference on Image Processing, Taipei, 2019. (査読有) (Accepted)

[2] Yu Wang and Jien Kato, "Good Choices for Deep Convolutional Feature Encoding", IEEE Winter Conference on Applications of Computer Vision 2019, Hawaii, 2019. (査読有)

[3] Longjiao Zhao, Yu Wang, Jien Kato and Hiromi Tanaka, "Good Practice of Using Deep Features in Content-based Image Retrieval", 25th International Workshop on Frontiers of Computer Vision, Gangneung, 2019. (査読有)

[4] Misaki Kodaira, Yu Wang, Jien Kato, Hiroshi Murase and Hiromi Tanaka, "Good Deep Features for Pedestrian Detection", 25th International Workshop on Frontiers of Computer Vision, Gangneung, 2019. (査読有)

[5] Dichao Liu, Yu Wang and Jien Kato, "Supervised Spatial Transformer Networks for Attention Learning in Fine-grained Action Recognition", International Conference on Computer Vision Theory and Applications 2019, Prague, 2019. (査読有)

- [6] Ying Ji, Yu Wang and Jien Kato, "A Fine-grained Dataset for Visual Violence Analysis", 画像の認識・理解シンポジウム 2018, 札幌コンベンションセンター, 2018. (査読無)
- [7] 小平美沙季, 王彧, 加藤ジェーン, "深層ローカル特徴を用いた歩行者検出", 画像の認識・理解シンポジウム 2018, 札幌コンベンションセンター, 2018. (査読無)
- [8] Longjiao Zhao, Yu Wang and Jien Kato, "Image Retrieval with Augmented Fine-tuned CNN Features", パターン認識・メディア理解研究会, 青山学院大学, 2018.3.18. (査読無)
- [9] Dichao Liu, Yu Wang and Jien Kato, "Hierarchical Multi-stream STNs for Fine-grained Action Recognition", パターン認識・メディア理解研究会, 慶応大学, 2017.12.16. (査読無)
- [10] Yu Wang and Jien Kato, "Collision Risk Rating of Traffic Scene from Dashboard Cameras", 2017 International Conference on Digital Image Computing: Techniques and Application, Sydney, 2017. (査読有)
- [11] Dichao Liu, Yu Wang and Jien Kato, "Evaluation of Triple-Stream Convolutional Networks for Action Recognition", 2017 International Conference on Digital Image Computing: Techniques and Application, Sydney, 2017. (査読有)
- [12] Longjiao Zhao, Yu Wang and Jien Kato, "Local Features Augmenting for Better Image Retrieval", 2017 International Conference on Digital Image Computing: Techniques and Application, Sydney, 2017. (査読有)
- [13] Dichao Liu, Yu Wang and Jien Kato, "Spatial Attention Selection for Fine-grained Action Recognition", 電気関係学会東海支部連合大会, 名古屋大学, 2017. (査読無)

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ等

無し

6. 研究組織

(1)研究分担者

無し

(2)研究協力者

無し

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。