

平成 30 年 6 月 20 日現在

機関番号：32613

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00159

研究課題名(和文)ディープラーニングを用いたビデオデータベースの自動アノテーションに関する研究

研究課題名(英文)Research on Automatic Annotation for Video Database Based on Deep Learning

研究代表者

陳 キュウ (CHEN, QIU)

工学院大学・情報学部(情報工学部)・准教授

研究者番号：00400292

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：近年、情報化社会の急速発展により、伝統的なテレビなどの視聴手段からオンライン動画視聴に変化しつつある。しかし、テレビで放送される映像や、ビデオカメラなどで撮影された映像などには、内容を説明したテキスト情報が存在しない。検索エンジンのように、キーワードなどを用いて簡単に意味に基づいた映像検索を実現するため、映像の内容を表すメタ情報をテキスト形式で付与する自動アノテーション手法が必要である。本研究では、ディープラーニングの方法を取り組み、ビデオの自動アノテーションを高精度に行う手法を提案し、従来手法により優れた性能を実現できた。

研究成果の概要(英文)：Due to the rapid development of information society, viewing means is being changed from traditional television to online video viewing in recent years. However, text information describing the contents does not exist in the videos broadcasted on television or shot with video cameras. In order to realize video retrieval just using keywords, automatic annotation method is necessary that gives meta information representing the contents of video in a text format. In this research, we propose a novel annotation algorithm for video database with high precision, which can realize superior performance compared with conventional methods.

研究分野：情報学

キーワード：アノテーション マルチメディア データベース ディープラーニング パターン認識 アルゴリズム
情報システム

1. 研究開始当初の背景

近年、ネットワーク接続速度の向上および大容量ストレージ装置の価格が下落に伴い、インターネット上で公開されるデジタルコンテンツの数が急速に膨らんでいる。伝統的なテレビなどの視聴手段からオンライン動画視聴に変化しつつある。例えば、動画サイト YouTube だけで、毎月 10 億人以上のユーザーが YouTube にアクセスし、毎月 60 億時間以上の動画が YouTube で視聴されている。さらに、1 分あたり約 100 時間分の動画が YouTube にアップロードされている。

しかし、テレビで放送される映像や、ビデオカメラなどで撮影された映像などには、内容を説明したテキスト情報が存在しないため、色や形などの画像特徴の類似性に基づいた検索になる。ビデオ映像はテキストに比べて情報量が非常に膨大な量になるため、処理に時間がかかるなどの問題が生じる。したがって、検索エンジンのように、キーワードなどを用いて簡単に意味に基づいた映像検索を実現するため、映像に対して、その内容を表すメタ情報をテキスト形式で付与する必要がある。このようなメタ情報の付与をビデオアノテーションと言う。

ビデオの自動アノテーション(図 1)とは現在の検索エンジンのように、キーワードなどを用いて簡単に意味に基づいた映像検索を実現するため、映像に対して、その内容を表すメタ情報をテキスト形式で付与することである。ビデオアノテーションでは、映像内のある時区間に対して、意味内容の記述が必要になる。記述される意味内容の基本としては、いつ (When)、どこで (Where)、だれが (Who)、どのように (How)、なぜ (Why) といった 5W1H に関する情報である。

現状では、ビデオコンテンツに対する Web 検索はいまだ実用化されているとは言い難い。ビデオコンテンツに対する検索にはさまざまな手法が存在するが、ビデオコンテンツを全自動で解析した結果に基づいて検索する場合、精度の観点からきわめて不十分である。検索の精度を十分に実用的なレベルに引き上げるためにはビデオコンテンツに検索や変換・編集等に有効な意味内容記述をな

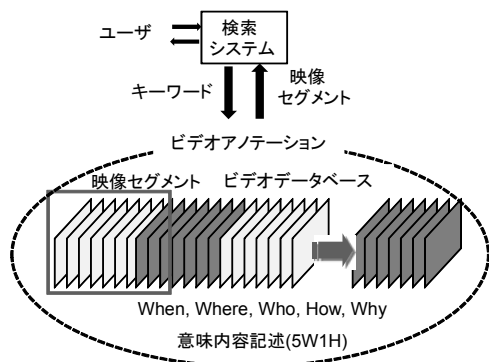


図 1. ビデオアノテーションの概念図

んらかの方法により付加する必要がある。また、コンピュータによるビデオコンテンツの自動解析を行い、人間がその解析結果を効率よく修正・補完できる半自動ビデオアノテーション手法が作成されているが、まだ不十分である。

申請者らはここまで「差分量子化によるビデオデータベースの自動アノテーションに関する研究」(基盤研究(C)-課題番号 24500104)において、差分量子化によるヒストグラム特徴を用いて、映像のショットを自動検出し、それから顔をキーとして、差分量子化を用いて顔検出、認識し、学習により自動的にビデオアノテーション手法を確立した。

また、「コードブック空間情報処理を用いた連想によるリアルタイム物体認識に関する研究」(若手研究(B)-代表者: 陳 キュウ)において、回転・スケール変化・照明変化に不変な SIFT 特徴の局所記述子を、ベクトル量子化ヒストグラム(VQ)を用いるアルゴリズムを提案した。従来の SIFT 局所記述子の特徴ベクトルは方向ヒストグラムを用いることに対して、ベクトル量子化(VQ)ヒストグラムを適用することにより、SIFT の識別性を向上させる VQ-SIFT 手法を提案した。

一方、ディープラーニング (Deep Learning) とは、従来よりも多くの層を持った人間の脳を模した構造をもつニューラルネットワークを用いる機械学習技術である。大きな特徴としては、多段に重ねることによって抽象的なデータの表現を獲得することができる。すでに画像認識、音声認識などの分野で、いろいろなベンチマークで従来技術を超える性能を出しており、脚光を浴びている。例えば、グーグルがディープラーニングを使って開発した人工知能「GoogLeNet」は、2014年8月に開催された画像認識技術のコンテスト「Imagenet Large Scale Visual Recognition Challenge 2014 (ILSVRC2014)」で首位となっている。

これらの研究成果は、上述ディープラーニングを用いたビデオの自動アノテーション技術実現に向けてその基礎となる重要な成果である。

2. 研究の目的

申請者らは「差分量子化によるビデオデータベースの自動アノテーションに関する研究」(基盤研究(C))において、差分量子化を用いて顔検出、認識し、学習により自動的にビデオアノテーション手法を確立した。この手法は顔をキーとして限定されているので、それ以外の種類のビデオに対してアノテーションにはまた別途対応する必要がある。本研究では、さらに多様な特徴を畳み込みニューラルネットワーク(CNN)により自動的に抽出し、総合判断でより広い意味でのビデオアノテーションができると考えられる。

本研究では、申請者らが提案した差分量子

化によるビデオデータベースの自動アノテーション技術に基づき、まずヒストグラム特徴を用いて、映像のショットを自動検出し、そして畳み込みニューラルネットワーク(CNN)を利用して、ビデオシーケンスの各フレームのシーンラベルを付与する。それから、時系列での特徴を抽出し、学習によりビデオデータベースの自動アノテーションを高精度かつ高速に実現する。

3. 研究の方法

本研究において、ディープラーニング(Deep Learning)を用いて、高速かつ高精度なビデオの自動アノテーションシステムを実現するため、フレーム画像のラベリング手法、ショットの認識手法、及び自動アノテーション手法からなるビデオデータベースの自動アノテーションアルゴリズムを提案する。

(1) フレーム画像のラベリング手法の開発

ビデオの各フレームにおいて、まずフレーム画像の局所領域を切り出し、それぞれ正規化を行う。それから、畳み込みニューラルネットワーク(CNN)を利用して、自動的にフレーム画像局所領域の特徴量を抽出し、サポートベクターマシン(SVM)でクラス識別し、フレーム内に存在する物体(人、テレビ、本など)を精度良く検出し、さらに、フレームにおける各ピクセルのフレームラベルを付与する。

(2) ショートの認識手法の開発

各フレーム画像内の物体検出及びラベリングを完成したあと、各フレームにおける大域的な動き特性を計算し、ピクセルレベルで複数フレームの情報を統合し、局所的な動き方向と速度を検出することにより、一連のフレームからなるショットの認識手法を確立する。また、顔検出されたビデオについて、高速かつ高精度で顔認識をし、ビデオの時間的連続性があることを利用し、得られた顔領域の中心位置を次の画像フレームにおける検索スタート位置とし、顔領域のヒストグラムをテンプレートとして使い、前のフレーム中の顔領域周辺を重点に置き、効率よくシーケンスの中で高速顔追跡し続けることによって、ショット認識を行う。

(3) 自動アノテーション手法の開発

ショット認識を行ったあとに、ディープラーニング手法によりビデオの自動アノテーションを高精度かつ高速に実現する。意味的にまとまりを持つ連続したショット列を畳み込みニューラルネットワーク(CNN)を利用して、制約付きボルツマンマシン(RBM)層を何段か繋げたディープボルツマンマシン(DBM)を使った学習手法によってビデオデータベースのアノテーション手法を確立する。

4. 研究成果

本研究の研究成果は以下のようになる。ディープラーニングによるビデオデータベースの自動アノテーションアルゴリズムを開発した。畳み込みニューラルネットワーク(CNN)を利用して、ビデオシーケンスの各フレームのシーンラベルを付与する。それから、時系列での特徴を抽出し、学習により高精度かつ高速にビデオデータベースの自動アノテーション手法を確立した。

映像のショットの自動検出には、申請者ら提案したヒストグラム特徴量を用いて、隣の画像フレーム同士頻度分布の変化を分析し、ディゾルブ・フェードなどを含むさまざまなタイプのビデオシーン切り替えを検出する。それからCNNを利用して、自動的にフレーム画像局所領域の特徴量を抽出し、サポートベクターマシン(SVM)でクラス識別し、フレーム内に存在するオブジェクトを検出し、さらに、フレームにおける各ピクセルのフレームラベルを付与する。その後、各フレームにおける大域的な動き特性を計算し、ピクセルレベルで複数フレームの情報を統合し、局所的な動き方向と速度を検出することにより、一連のフレームからなるショットを認識した。

また、新たに提案する「方向ベクトル特徴量」を用いた高速探索を試みた。方向ベクトル特徴量を既存の特徴量と組み合わせることで映像の探索精度が向上することを示した。また、時系列アクティブ探索の適用により、無駄な照合を削減することで探索の高速化ができた。

そして、本研究で行ったシーンの識別手法について、新たにスパースオートエンコーダに基づいた手法を提案した。概要として、フレーム画像から局所的なHOG特徴を抽出し、スパース・オートエンコーダー(sparse autoencoder)によりエンコーディングし、得られたスパース特徴をピラミッドプーリング、及び局所正規化することによってフレーム画像を表現し、より精度の高いシーンの識別を実現できた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

- (1) Lin Xie, Feifei Lee, Li Liu, Zhong Yin, Yan Yan, Weidong Wang, Junjie Zhao, and Qiu Chen, "Improved Spatial Pyramid Matching for Scene Recognition", Pattern Recognition, Vol. 82, pp. 118-129, 2018. (査読有)
- (2) Yan Yan, Feifei Lee, Xueqian Wu, Qiu Chen, "Face Recognition Algorithm Using Extended Vector Quantization Histogram Features", Plos One, Vol. 13, Issue. 1, e0190378, 2018. (査読有)

- (3) Qiu Chen, Koji Kotani, Feifei Lee, and Tadahiro Ohmi, “An Improved Face Recognition Algorithm Using Histogram-based Features in Spatial and Frequency Domains”, *Int'l Journal of Computer, Electrical, Automation, Control and Information Engineering*, Vol. 10, No. 2, pp. 385-389, Apr. 2016. (査読有)
- (4) Qiu Chen, Koji Kotani, Feifei Lee, and Tadahiro Ohmi, “An Improved Histogram-based Features in Low-frequency DCT Domain for Face Recognition”, *International Journal of Machine Learning and Computing*, Vol. 5, No.5, pp. 404-408, Oct. 2015. (査読有)
- [学会発表] (計 17 件)
- (1) 綱島 秀樹, 佐藤 祥, 星 泰成, 陳 キュウ, “ディープラーニングによるモバイル顔認証システム”, 2018 電子情報通信学会総合大会講演論文集, D-12-56, 2018 年 3 月. (査読無)
- (2) 中村 充志, 佐藤 祥, 陳 キュウ, “画像の感性を反映させたフォントの自動生成手法”, 2018 電子情報通信学会総合大会講演論文集, D-12-57, 2018 年 3 月. (査読無)
- (3) 星 泰成, 佐藤 祥, 綱島 秀樹, 陳 キュウ, “敵対的生成ネットワークを用いた顔画像生成手法”, 2018 電子情報通信学会総合大会講演論文集, D-12-61, 2018 年 3 月. (査読無)
- (4) 森野 光輝, 佐藤 祥, 陳 キュウ, “異なる空間情報を利用した類似画像検索手法”, 2018 電子情報通信学会総合大会講演論文集, D-12-61, 2018 年 3 月. (査読無)
- (5) Feifei Lee, Junjie Zhao, Koji Kotani, and Qiu Chen, “Video Copy Detection Using Histogram Based Spatio-temporal Features”, 10th Int'l Congress on Image and Signal Processing, Biomedical Engineering and Informatics, Shanghai, China, Oct. 2017. (査読有)
- (6) Qiu Chen, Feifei Lee, Koji Kotani, “Improved Multiple Histogram-based Features for Robust Face Recognition”, 10th Int'l Congress on Image and Signal Processing, Biomedical Engineering and Informatics, Shanghai, China, Oct. 2017. (査読有)
- (7) 宮本朋希, 長田唯輝, 陳 キュウ, “複数ヒストグラム特徴量の統合によるビデオデータベースの高速探索-特徴量及び探索手法の検討”, 2017 電子情報通信学会総合大会講演論文集, D-12-32, 2017 年 3 月. (査読無)
- (8) 長田唯輝, 宮本朋希, 陳 キュウ, “複数ヒストグラム特徴量の統合によるビデオデータベースの高速探索-探索システム性能の評価”, 2017 電子情報通信学会総合大会講演論文集, D-12-29, 2017 年 3 月. (査読無)
- (9) 佐藤 祥, 陳 キュウ, “データ量削減による画像検索の高速化に関する検討”, 2017 電子情報通信学会総合大会講演論文集, D-12-30, 2017 年 3 月. (査読無)
- (10) 坂本晃市, 佐藤 祥, 陳 キュウ, “実空間及び周波数空間特徴量を用いた類似画像検索手法の検討”, 2017 電子情報通信学会総合大会講演論文集, D-12-31, 2017 年 3 月. (査読無)
- (11) 宮本 朋希, 長田 唯輝, 陳 キュウ, “複数ヒストグラム特徴量の統合によるビデオデータベース高速探索の精度向上に関する検討”, 信学技報, Vol. 116, No. 463, IE2016-110, pp. 247-251, Feb. 2017. (査読無)
- (12) Qiu Chen, Koji Kotani, and Feifei Lee, “Face Recognition Using Multiple Histogram Features in Spatial and Frequency Domains”, *Proceeding of 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2016)*, pp. 204-208, Naples, Italy, Nov. 2016. (査読有)
- (13) Sho Sato, Kazuhiro Kobayashi, and Qiu Chen, “Face Verification System for Android Mobile Devices Using Histogram Based Features”, *Proceeding of the 1st International Workshop on Pattern Recognition (IWPR 2016)*, pp. 35-41, Tokyo, Japan, May, 2016. (査読有)
- (14) 藤井祐輔, 小林積博, 佐藤 祥, 陳 キュウ, “ヒストグラム特徴量を用いた画像検索手法の検討”, 2016 電子情報通信学会総合大会講演論文集, D-12-94, pp.163-163, 2016 年 3 月. (査読無)
- (15) 鏡 翔太, 山部未葵, 陳 キュウ, “ヒストグラム特徴を用いたビデオデータベースの高速探索”, 2016 電子情報通信学会総合大会講演論文集, D-12-95, pp.164-164, 2016 年 3 月. (査読無)
- (16) Qiu Chen, Koji Kotani, Feifei Lee, and Tadahiro Ohmi, “A Fast Search Algorithm for Large Video Database Using HOG Based Features”, *Proceeding of the 5th International Conference on Signal and*

Image Processing (SIP 2016), pp. 35-41,
Sydney, Australia, Feb. 2016. (査読有)

- (17) Kazuhiro Kobayashi, Qiu Chen, “Image Retrieval Using Features in Spatial and Frequency Domains Based on Block-Division”, Proceeding of 2015 International Conference on Computational Science and Computational Intelligence, pp. 448-453, Las Vegas, USA, Dec. 7-9, 2015. (査読有)

6. 研究組織

(1) 研究代表者

陳 キュウ (CHEN, Qiu)
工学院大学・情報学部・准教授
研究者番号：00400292

(2) 研究分担者

小谷 光司 (KOTANI, Koji)
秋田県立大学・システム科学技術学部・
教授
研究者番号：20250699