

令和 2 年 5 月 28 日現在

機関番号：34416

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K12687

研究課題名（和文）映像シーン検索を目的としたコーパスの自動構築と時系列アノテーション法の開発

研究課題名（英文）Development of methods for video corpus construction and time series annotation for video scene retrieval

研究代表者

吉田 壮 (Yoshida, Soh)

関西大学・システム理工学部・助教

研究者番号：70780584

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）：本研究では、ウェブから学習データを自動収集して構築したコーパスを基に、映像アノテーションおよび映像シーン検索する手法を開発した。ここでは、コーパス上のラベルについてもウェブデータ集合の解析から自動抽出する。さらに、ラベルごとの学習データセットの信頼度を向上させるためにキーワード検索の精度を高精度化するリランキング手法の開発を行った。映像シーン検索には、ユーザが望むシーンを含む映像を例示する適合性フィードバックを導入した検索システムを構築した。以上の有効性は、実際にYouTubeを対象とした実験データセットを構築して確認した。

研究成果の学術的意義や社会的意義

これまでの映像アノテーションに関する研究は、その内容に応じて専門家がラベルを定義し、手で収集された映像コーパスを前提としており、アノテーションを行うために必要なラベル及び学習データセットを自動的に取得することが困難であった。それに対して、本研究では、映像内容と関連するラベルを推定し、従来実現困難であった自動構築された映像コーパスを基にアノテーションを行う点、そして、その結果から類似シーンを関連付けて映像検索結果を提示する点が学術的に意義がある。ユーザが望む映像を的確に提示する技術であり、映像検索の効率化が見込め、インターネットを通じて映像を配信するサービスと柔軟に連携されることが期待できる。

研究成果の概要（英文）：We developed methods for supervised video annotation and video scene retrieval based on a corpus constructed from automatically collected learning data from the web. The labels on the corpus are also automatically extracted from the analysis of the web dataset. Furthermore, we developed a reranking method to improve the accuracy of keyword searches to improve the reliability of the training dataset. For developing a video scene retrieval algorithm, we use relevance feedback, which the user can interact with by specifying some example videos from the initial search results. The effectiveness of the proposed methods was confirmed by conducting experiments on the YouTube dataset.

研究分野：映像検索

キーワード：映像検索 映像認識

様式 C - 19, F - 19 - 1, Z - 19 (共通)

1. 研究開始当初の背景

(1) スマートフォンやタブレット端末の普及に伴い、インターネットによる映像配信サービスの利用者数が増加している。具体的な数値としては、YouTube が公表した統計データでは、映像を視聴する利用者数はひと月平均で 10 億人を超え、利用者の視聴時間のひと月の総計は 60 億時間に昇る。このような背景の下、ユーザが望む映像を検索するための技術開発が急務である。

(2) 映像配信サービスが提供する映像には検索を容易にするため、タイトル、コメント、及びタグなどのメタデータが付与されている。しかしながら、付与されるメタデータは映像全体に対して共通かつ代表的なものに留まり、それらが指す時間区間が不明確である。一方、これまで国内外において、自動でメタデータを付与する従来手法は、内容に応じて専門家が定義した学習データセット(映像コーパス)を前提とした統計的機械学習を用いることから自動化のために必要なラベルとそれらに対応するデータセットの整備が困難である。また、近年、機械学習用コーパスの情報源として、SNS にアップロードされたマルチメディアデータを用いたコーパスの構築法が提案されているがその自動化には至っていない。

(3) そこで本研究では、以下の 2 点を解決すべき課題と考えた。

【課題 1】映像全体ではなくシーンへ適用範囲を広げる時系列アノテーションを行うためには、映像内容を詳細に観察する必要がある。一方で、未定義のラベルを映像に付与することは困難である。つまり、未知のラベルを推定する手法の構築が必要である。

【課題 2】ウェブは対象とする内容以外のデータが多く含まれる雑多な情報源である。従来研究は依然として手動によるフィルタリングを行っていることから、膨大な量のデータを扱うことが困難である。つまり、雑多な情報源から対象とする撮像内容を含むデータを高精度に取り出す手法の構築が必要である。

上記二つの課題を解決することにより、映像コーパスの自動構築が可能となり、それを利用した識別モデルの学習によって時系列アノテーションを実現することが可能となる。

2. 研究の目的

研究の背景(3)で挙げた 2 つの課題を以下のように解決することを考える。

【課題 1 の解決】アノテーションの対象とする映像に付与されたメタデータを基に検索キーワードを生成して SNS に与えることにより、画像及び映像などのマルチメディアデータを収集する。このとき、収集したデータと付随するテキスト情報を統合的に処理するマイニングから関連ラベルの抽出手法を構築する。

【課題 2 の解決】データの精練を行うため、異種のデータ間で関連性を求める手法を用いる。具体的には、これまでの研究で導出した「対象物体である可能性を表す確率」を応用してラベル付き学習データベースの自動作成、つまり、映像コーパスを自動的に構築する。

本研究では、映像コーパスの自動構築を可能とすることで、映像の時系列アノテーションの自動化の実現を目指す。具体的には

- (1) マルチメディアデータを活用した関連ラベルの抽出手法の開発
- (2) SNS から収集したデータの精練と映像コーパスの自動構築法の開発
- (3) 映像の時系列アノテーション法の開発
- (4) 映像シーン検索法の開発

までを実施することなどが目的である。

3. 研究の方法

まず、映像コーパスの自動構築法の開発に取り組む。具体的には、映像に付随するメタデータを基に、マルチメディアデータを SNS から収集し、それを情報源として用いたマイニングから撮像内容と関連するラベル(関連ラベル)を推定する手法を開発する。次に、同一ラベルの内容を含むマルチメディアデータをクラスタリングすることで新たな学習データセット(映像コーパス)を構築する手法を開発する。続いて、得られた映像コーパスから各ラベル内容を識別するモデルを構築する。識別モデルの出力値を観測することで時系列アノテーションを行い、シーンにラベル付けを施す。このとき、ラベルのテキスト情報とシーンの映像特徴の両者を基に類似シーン同士を関連付けて検索結果を可視化することにより、望む映像を素早く探し出すことを可能とする提示方法を検討する。

4. 研究成果

本研究課題では、下記のような研究成果を得た。

- (1) マルチメディアデータを活用した関連ラベルの抽出手法の開発

ここでは、関連ラベルの抽出のため、映像特徴とテキスト特徴の利用を検討した。映像の部分集合を映像特徴に基づいたクラスタリングを行い、映像内容ごとに事前分割されたクラスタから代表するラベルを抽出する手法を開発した。具体的には、映像に付随するラベルを、分散表現を用いてベクトル化し、ベクトルの中心性を算出する。そして、各クラスタにおいて最も中心性の高いラベルを抽出する。このとき、クラスタ純度の高いラベルを判断する処理を導入し、2 段階でラベルを選出することで、関連性の低いラベルが抽出されることを抑制した。以上により、ある 1 つのラベル、関連する複数のラベルを自動抽出することを可能とした。

- (2) SNS から収集したデータの精練と映像コーパスの自動構築法の開発

(1)の手法で抽出されたラベルをキーワードに用いた検索から映像を収集する．これにより，ラベルと映像集合がセットとなった映像コーパスを自動構築する．雑多なソーシャルメディアを精練するため，キーワードと実際の映像の意味内容の間で生じるセマンティックギャップを抑制する手法を開発した．具体的には，キーワード検索で得た結果を映像特徴の整合性を最大化するように検索順位を並べ替えるリランキング手法を開発した．実際に有効性の検証は YouTube を用いた映像検索で行った．15 個のラベルを用いて検索した上位 100 件の精度は，キーワード検索と比較すると平均して 20%の精度向上を実現した．開発した本手法の適用により，ラベルと映像の視覚的内容の不一致を防ぐことを可能とし，映像コーパスの信頼度を向上するとともにコーパス作成時のユーザの負担を軽減することが期待できる．

(3) 映像の時系列アノテーション法の開発

アノテーションのために，Convolutional Neural Network (CNN)の適用を検討した．(2)の手法の適用でデータの精練を可能とした一方で，ラベルと内容が一致しない映像を取り切れない場合がある．そこで，誤ラベルが存在する学習データセットから学習した場合でも頑健な CNN モデルを検討した．具体的には，同じクラス同士のサンプルの類似度は高いことを仮定して，CNN の学習中に中間層から得られる特徴表現の変化に基づいて重みの更新を適応的に変化させる手法を提案した．本手法の開発を通じて，CNN など高精度で知られるアノテーション手法の学習に，誤ラベルの存在を許容することを可能とする．それとともに，ソーシャルメディアを学習データとして活用する方式の導入が期待できる．本手法は映像をフレーム分割した後の適用を想定しているが，今後は，エンドツーエンドで映像の時系列にアノテーションを行う手法について検討していきたいと考えている．

(4) 映像シーン検索法の開発

ここでは，ユーザが望むシーンを含む映像を検索するために，適合性フィードバックを導入した検索システムを提案した．適合性フィードバックとは，検索結果の適合および不適合をユーザ本人が例示することで，検索精度を改善する枠組みである．この枠組みを利用してキーワード検索で得られた映像から検索意図に適したシーンを含む映像をユーザが複数件選択し，それらの映像との類似度からランキングの適合性を評価する手法を提案した．映像データセット YouTube-8M を用いた実験で手法を評価し，シーン内容を表すキーワードを用いた検索と比較して，上位 10 件の検索精度が 22.2%の向上を果たせた．これは，多くのユーザが携帯端末を使用して映像を閲覧していることを考慮すると，上位 10 件の映像の精度向上を大きく示したことから提案手法は有効であるといえる．これらの結果から，映像のシーン検索の目処が立ったことから，(1)～(3)の手法でアノテーションされた映像を用いて，時間情報とともに望むシーンを直接検索可能とする手法の可能性について検証していきたいと考えている．

5. 主な発表論文等

〔雑誌論文〕 計2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 1件）

1. 著者名 YOSHIDA Soh, OGAWA Takahiro, HASEYAMA Miki, MUNAYASU Mitsuji	4. 巻 E101.D
2. 論文標題 Graph-Based Video Search Reranking with Local and Global Consistency Analysis	5. 発行年 2018年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 1430 ~ 1440
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1587/transinf.2017EDP7277	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 FUJII Takamasa, YOSHIDA Soh, MUNAYASU Mitsuji	4. 巻 E102.A
2. 論文標題 Video Search Reranking with Relevance Feedback Using Visual and Textual Similarities	5. 発行年 2019年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 1900 ~ 1909
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1587/transfun.E102.A.1900	査読の有無 有
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 5件）

1. 発表者名 武田 啓志, 吉田 壮, 棟安 実治
2. 発表標題 誤ラベルを含む教師データを用いたCNNによる画像認識
3. 学会等名 電子情報通信学会スマートインフォメディアシステム研究会
4. 発表年 2019年

1. 発表者名 S. Yoshida and M. Muneyasu
2. 発表標題 A Graph-based Video Visual Reranking Method via Heterogenous Graph Analysis
3. 学会等名 2019 International Workshop on Smart Info-Media Systems in Asia (国際学会)
4. 発表年 2019年

1. 発表者名 H. Takeda, S. Yoshida and M. Muneyasu
2. 発表標題 Tag-based Video Retrieval with Social Tag Relevance Learning
3. 学会等名 2019 IEEE 8th Global Conference on Consumer Electronics (国際学会)
4. 発表年 2019年

1. 発表者名 T. Fujii, S. Yoshida, and M. Muneyasu
2. 発表標題 Video Retrieval by Reranking and Relevance Feedback with Tag-Based Similarity
3. 学会等名 2018 IEEE 7th Global Conference on Consumer Electronics (国際学会)
4. 発表年 2018年

1. 発表者名 S. Yoshida, and M. Muneyasu
2. 発表標題 Video Search Reranking on Multi-Layer Graphs Based on Combination of Video Features Using Subspace Analysis
3. 学会等名 2018 International Workshop on Smart Info-Media Systems in Asia (国際学会)
4. 発表年 2018年

1. 発表者名 吉田 壮, 棟安 実治
2. 発表標題 コンテンツの類似度に基づいた動画共有サイトで付与されたタグのランキング手法
3. 学会等名 電子情報通信学会スマートインフォメディアシステム研究会
4. 発表年 2018年

1. 発表者名 藤井 孝匡, 吉田 壮, 棟安 実治
2. 発表標題 マルチタグによるコンテンツ表現を考慮した映像のリランキング手法
3. 学会等名 電子情報通信学会スマートインフォメディアシステム研究会
4. 発表年 2018年

1. 発表者名 藤井 孝匡, 吉田 壮, 棟安 実治
2. 発表標題 タグベクトル表現を用いたWeb映像検索結果のリランキング
3. 学会等名 スマートインフォメディアシステム研究会
4. 発表年 2018年

1. 発表者名 吉田 壮, 棟安 実治
2. 発表標題 異種メタデータを用いたグラフベースソフトクラスタリングに基づくWeb映像リランキング
3. 学会等名 電子情報通信学会スマートインフォメディアシステム研究会
4. 発表年 2018年

1. 発表者名 T. Fujii, S. Yoshida, and M. Muneyasu
2. 発表標題 Feedback Assisted Multi-Modality Reranking for Web Video Search
3. 学会等名 2017 International Workshop on Smart Info-Media Systems in Asia (SISA 2017) (国際学会)
4. 発表年 2017年

1. 発表者名 吉田 壮, 小川 貴弘, 長谷山美紀, 棟安 実治
2. 発表標題 Web Video Search Reranking using a Heterogeneous Graph-Based Soft Clustering Approach
3. 学会等名 画像の認識・理解シンポジウム (MIRU2017)
4. 発表年 2017年

1. 発表者名 藤井 孝匡, 吉田 壮, 棟安 実治
2. 発表標題 ユーザフィードバックを利用したコンテンツの適合性に基づく映像検索
3. 学会等名 電子情報通信学会ソサイエティ大会,
4. 発表年 2017年

1. 発表者名 吉田 壮, 中村健太郎, 棟安 実治
2. 発表標題 マルチレイヤグラフを用いた映像特徴の統合に基づく映像検索手法の精度向上
3. 学会等名 電気学会研究会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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