

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：62615

研究種目：基盤研究(B)

研究期間：2011～2013

課題番号：23300041

研究課題名(和文)大規模マルチメディアアーカイブにおける潜在的リンク解析に基づくマイニング

研究課題名(英文)Mining based on latent link analysis for large-scale multimedia archive

研究代表者

佐藤 真一 (Sato, Shin'ichi)

国立情報学研究所・コンテンツ科学研究系・教授

研究者番号：90249938

交付決定額(研究期間全体)：(直接経費) 14,600,000円、(間接経費) 4,380,000円

研究成果の概要(和文)：大規模放送映像アーカイブ、Web上の画像・映像データ等の大量のマルチメディアアーカイブ中に頻出する、物体等に相当する部分画像を検出・同定する手法の実現を目指す。また、こうして得られる、同定された部分画像間の関連を潜在的リンクとみなし、アーカイブから得られる潜在的リンクにより構成されるグラフ構造に対するマイニング技術について検討を行う。本成果により、マルチメディアアーカイブ中の特定の物体の出現の列挙ならびにそれに基づく統計解析、アノテーション等の他のメタデータとの共起解析等による物体自動名前付け、物体認識システムの完全自動生成等が可能となる。

研究成果の概要(英文)：We address partial image detection and identification which may correspond to recurrent objects in huge amount of multimedia archives such as large-scale broadcast video archive and Web images and videos. We regard detected and identified partial images as latent link and study mining techniques for graph structure composed of latent links obtained from archives. The outcome of this project enables enumeration of specific objects appear in multimedia archive and their statistical analysis, object auto-annotation based on co-occurrence analysis with metadata and annotation, self-taught object recognition system, and so on. Partial image detection and identification, as the most important technical component, will be implemented by extending our former research outcome of image/video copy detection and mining techniques.

研究分野：総合領域

科研費の分科・細目：情報学 メディア情報学・データベース

キーワード：マルチメディアアーカイブ 潜在的リンク解析 部分画像同定 映像マイニング

1. 研究開始当初の背景

画像・映像の意味内容解析は、特徴点群の局所特徴空間中の分布に基づく手法、すなわち **Bag of Keypoint(BoK)** に基づく手法が主流である。この方法は、画像・映像アーカイブから得た大量の特徴点から **SIFT** 等の局所特徴量を算出し、**k-means** 等のクラスタリングにより離散化することで「単語」とみなし、単語の頻度分布に基づいて画像・映像の内容解析(画像・映像の概念クラスへの分類など)を行う手法である。しかし、点では単語に類するような意味情報は記述できないため、記述として本質的に弱いという問題がある。一方、点から拡張しようというアプローチもある。特徴点の対や三つ組を用いようというアプローチ、あるいは物体や部品レベルの記述となることを期待して画像セグメントを処理単位として用いようというアプローチもあるが、いずれも意味分類精度の向上には成功していない。

特徴点を越えた、特徴点群に基づく照合に関する研究では、画像・映像コピー検出分野が最も研究が進んでいる。そもそもは商業検出や違法画像・映像検知を目的として研究が進んだ。**BoK** のような特徴点の特徴空間内の分布のみならず、複数の対応特徴点間の幾何的整合性検証によるより高精度な照合の実現、部分領域への対応等のきわめて精度の高い照合、大規模アーカイブを扱う高速・スケーラブルな手法、検索・検出のみならず、アーカイブ中のすべてのコピーを列挙するマイニング技術についても効率のよい手法が実現されている。

われわれも、映像コピー検出とマイニング技術、ならびにそれに基づく映像アーカイブ構造化技術について基盤研究(B)「同一映像断片照合による放送映像アーカイブの構造化」(平成 20-22 年度)として検討している。こうした成果から示されているとおり、一定の研究成果が上がったが、一方ショットレベルのコピー検出では粒度が荒く、またアーカイブ中の頻度も少ないという問題も明らかになった。物体レベルのコピー検出を目指した部分領域コピー検出並びにマイニングにより、こうした問題が大幅に改善されると考えた。

2. 研究の目的

大規模放送映像アーカイブ、Web 上の画像・映像データ等の大量のマルチメディアアーカイブ中に頻出する、物体等に相当する部分画像を検出・同定する手法の実現を目指す。また、こうして得られる、同定された部分画像間の関連を潜在的リンクとみなし、アーカイブから得られる潜在的リンクにより構成されるグラフ構造に対するマイニング技術について検討を行う。本成果により、マルチメディアアーカイブ中の特定の物体の出現の列挙ならびにそれに基づく統計解析、アンテーション等の他のメタデータとの共起解析等による物体自動名前付け、物体認識シス

テムの完全自動生成等が可能となる。最も重要な要素技術である部分画像検出・同定は、実現済みの映像・画像コピー検出・マイニング技術を、部分領域対応に拡張することで実現する。

3. 研究の方法

本研究では、以下の研究項目について具体的な検討を行った。

(1) 部分領域照合技術と画像識別

部分領域として画像中の矩形領域を想定し、その中での **BoK** 表現に基づく照合を実現した。また、この技術を応用し、画像を単位とした物体認識用ラベルに基づき、当該物体に対応する部分領域を自動的に推定し、高精度に画像識別を行う技術を実現した。

(2) コマーシャル映像マイニングの大規模化
長時間の映像ストリーム中で繰り返し現れる映像部分区間をコマーシャル映像として自動検出・同定するコマーシャル映像マイニング技術はすでに実現済みであったが、これを拡張し、複数のマイニング結果を適切に統合する方法を新たに開発した。これにより、原理的には制限なく大きな映像アーカイブからのコマーシャル映像マイニングが実現可能となった。

(3) **BoK** に基づく高精度部分領域照合技術

BoK に基づく方法ながら、これまでになく高精度の部分領域照合が可能となる手法を実現した。これにより、次の二つの項目における物体レベルの検索や照合が可能となった。

(4) コマーシャル映像マイニングに基づく放送戦略解析システム

コマーシャル映像マイニングの成果に基づき、その放送パターンの可視化に基づく放送戦略の解析、適切なコマーシャルの事例画像もしくはキーワードによる検索等の技術を統合した、コマーシャル映像マイニングに基づく放送戦略解析システムを実現した。

(5) 放送映像と Web 情報との融合による影響解析システム

放送映像アーカイブと、**blog** から抽出した画像との間の部分領域もしくは共起物体に基づく関連性を検出し、これらの時間的変動パターンなどを解析することにより、相互の影響解析を実現した。

4. 研究成果

(1) 部分領域照合技術と画像識別

画像識別問題において、識別ラベルは画像内に写っている事物に基づき付与されるが、通常ラベルは画像に対して付与されるものの、対象物体は画像内の一部のみを占めている。われわれは、画像を単位としたラベルに基づき、物体領域に対応する部分領域を推定しつつ、より高精度な画像識別を行う技術を実現した。

ここで、部分領域としては矩形領域を用い、**BoK** を特徴量として用い、矩形に対応する特徴量をインスタンス、画像全体をバッグとみ

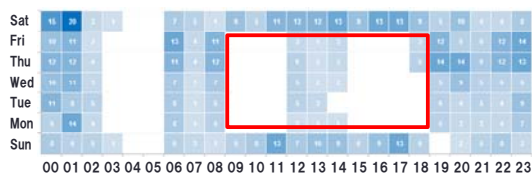
なし、Multiple Instance Learning (MIL) アルゴリズムを用いて解決している。特に、インスタンスが矩形領域であり、インスタンス間に画像中での領域間の近さが重要な情報であることから、これを考慮に入れた MIL アルゴリズムを新たに考案した。これにより、従来よりも高い識別性能を達成することに成功した。本成果については[12][14]にまとめている。

(2) コマーシャル映像マイニングの大規模化
長時間の映像ストリーム中で繰り返し現れる映像部分区間をコマーシャル映像として自動検出・同定するコマーシャル映像マイニング技術はすでに実現済みであったが[16]、より大規模(数年規模)の映像アーカイブからコマーシャル映像を検出・同定するため、数か月単位程度のアーカイブを独立に処理し、それぞれから出力されるコマーシャル映像のクラスタを高品質に統合する方法を開発した。新たに開発した bag of fingerprint 法では、各コマーシャル映像から抽出する fingerprint に基づき、クラスタを bag of fingerprint として表現し、これらの間の類似度により、クラスタの統合を行っている。これにより、原理的には制限なく大きな映像アーカイブからのコマーシャル映像マイニングが実現可能となった。本手法は[13]にまとめている。

(3) BoK に基づく高精度部分領域照合技術
BoK に基づく方法ながら、これまでにない高精度の部分領域照合を可能とする方法について検討を行った。これらは、100 万次元という極めて疎なコードブックに基づく方法[10]と、問い合わせ画像とデータベース画像との間にある非対称性を考慮した新たな照合尺度を利用する方法[6]とであり、これらによりきわめて高精度かつ高速な部分領域照合が可能となった。

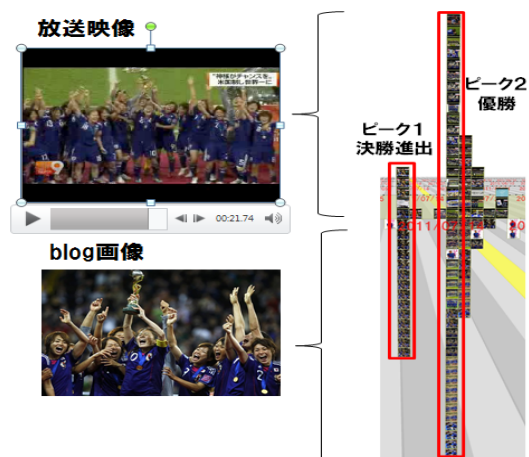
(4) コマーシャル映像マイニングに基づく放送戦略解析システム

コマーシャル映像マイニングにより、特定のコマーシャル映像がどの曜日・どの時間帯に放送されているのかなどの解析が可能である。そこで、この結果を可視化することにより、放送戦略の解析を実現した。また、適切なコマーシャルの事例画像もしくはキーワードによる検索等の技術も統合し、コマーシャル映像マイニングに基づく放送戦略解析システムを実現した。この成果は[2][5]にまとめている。また、可視化結果を以下に示す。



(5) 放送映像と Web 情報との融合による影響解析システム
二種類のデータベース(より具体的には放送

映像アーカイブと Web 上の画像データ)との間の部分領域マイニングを実現した。部分領域マイニングとしては、部分領域照合に基づく join 処理であり、より具体的には、Web から取得した 46000 画像と、放送映像 6000 時間分とで、相互に同一物体を含む画像を検出した。この結果、Web と放送映像との影響解析が可能となり、双方のメディアとしての特性の違いが観測できた。本成果は[4]にて公表している。また、可視化結果を以下に示す。



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 3 件)

- [1] Thanh Duc Ngo, Hung Thanh Vu, Duy-Dinh Le, and Shin'ichi Satoh, "Face Retrieval in Large-Scale News Video Datasets," IEICE Trans. on Information and Systems, Vol. E96-D, No.8, pp.1811-1825, 2013.
- [2] Xiaomeng Wu and Shin'ichi Satoh, "Ultrahigh-Speed TV Commercial Detection, Extraction, and Matching," IEEE Trans. on Circuits and Systems for Video Technology, Vol.23, No.6, pp.1054-1069, 2013.
- [3] Shin'ichi Satoh, "Simple Low-Dimensional Features Approximating NCC-based Image Matching," Pattern Recognition Letters, Vol. 32, Issue 14, pp.1902-1911, 2011.

[学会発表] (計 13 件)

- [4] Masahiko Itoh, Masashi Toyoda, Cai-Zhi Zhu, Shin'ichi Satoh, and Masaru Kitsuregawa, "Image Flows Visualization for Inter-Media Comparison," IEEE PacificVis 2014, pp.129-136, Yokohama, Japan, March 4-7, 2014.
- [5] Cai-Zhi Zhu, Siriwat Kasamwattanakote, Xiaomeng Wu,

- and Shin'ichi Satoh, "Tell me about TV commercials of this product," The International Conference on MultiMedia Modeling (MMM2014), Dublin, Ireland, Jan. 8-10, 2014.
- [6] Cai-Zhi Zhu, Hervé Jégou, and Shin'ichi Satoh, "Query-Adaptive Asymmetrical Dissimilarities for Visual Object Retrieval," International Conference on Computer Vision (ICCV2013), Sydney, Dec. 3-6, 2013.
- [7] Cai-Zhi Zhu, Xiao Zhou, Shin'ichi Satoh, "Bag-of-Words against Nearest-Neighbor search for visual object retrieval," Asian Conference on Pattern Recognition (ACPR2013), Naha, Japan, Dec. 6-8, 2013.
- [8] Cai-Zhi Zhu and Shin'ichi Satoh, "Evaluation of Visual Object Retrieval Datasets," International Conference on Image Processing (ICIP2013), Melbourne, Australia, Sep. 15-18, 2013.
- [9] Duy-Dinh Le and Shin'ichi Satoh, "Auto Face Re-Ranking By Mining the Web and Video Archives," Proc. of Computer Vision and Pattern Recognition (CVPR), Providence, USA, Jun. 16-21, 2012.
- [10] Cai-Zhi Zhu and Shin'ichi Satoh, "Large Vocabulary Quantization for Searching Instances from Videos," Proc. of International Conference on Multimedia Retrieval (ICMR2012), Hong Kong, China, Jun. 5-8, 2012.
- [11] Duy-Dinh Le and Shin'ichi Satoh, "Summarizing Large News Video Archives by Event Ranking," International Conference on Semantic Computing (ICSC2011), pp. 228-234, Stanford University, Palo Alto, USA, Sep. 18-21, 2011.
- [12] Thanh Duc Ngo, Duy-Dinh Le, and Shin'ichi Satoh, "Improving Image Categorization by Using Multiple Instance Learning with Spatial Relation," International Conference on Image Analysis and Processing (ICIAP 2011), Ravenna, Italy, Sep. 14-16, 2011.
- [13] Xiaomeng Wu and Shin'ichi Satoh, "Commercial Mining Based on Temporal Recurrence Hashing Algorithm and Bag-Of-Fingerprints Model," International Conference on Image Processing (ICIP2011), Brussels, Belgium, Sep. 11-14, 2011.
- [14] Thanh Duc Ngo, Duy-Dinh Le, and Shin'ichi Satoh, "Boosting global scene classification accuracy by

discriminative region localization," International Conference on Image Processing (ICIP2011), Brussels, Belgium, Sep. 11-14, 2011.

- [15] Hung Thanh Vu, Thanh Duc Ngo, Thao Ngoc Nguyen, Duy-Dinh Le, Shin'ichi Satoh, Bac Hoai Le, and Duc Anh Duong, "Fast Face Sequence Matching in Large-scale Video Databases," International Conference on Image Processing (ICIP2011), Brussels, Belgium, Sep. 11-14, 2011.
- [16] Xiaomeng Wu and Shin'ichi Satoh, "Temporal Recurrence Hashing Algorithm for Mining Commercials from Multimedia Streams," International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), Prague, Czech, May 22-27, 2011.

〔図書〕 (計 1 件)

- [17] Xiaomeng Wu, Sebastien Poullot, and Shin'ichi Satoh, "Multimedia Duplicate Mining towards Knowledge Discovery," in Frank Y. Shih ed., "Multimedia Security: Watermarking, Steganography, and Forensics" Taylor & Francis Group, LLC, 2012.

〔産業財産権〕 (計 0 件)

なし

〔その他〕 (計 0 件)

なし

6. 研究組織

(1) 研究代表者

佐藤 真一 (SATOH Shin'ichi)

国立情報学研究所・コンテンツ科学研究系・教授

研究者番号：90249938

(2) 研究分担者

なし

(3) 連携研究者

レ ユイディン (Le, Duy-Dinh)

国立情報学研究所・コンテンツ科学研究系・准教授

研究者番号：80450170

武 小萌 (Wu, Xiaomeng)

国立情報学研究所・コンテンツ科学研究系・特任研究員

研究者番号：20462179