

機関番号：62615

研究種目：基盤研究(B)

研究期間：2008 年 ～ 2010 年

課題番号：20300041

研究課題名（和文） 同一映像断片照合による放送映像アーカイブの構造化

研究課題名（英文） Structuring Broadcast Video Archive based on Video Near-Duplicate Detection

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

研究成果の概要（和文）：

本研究では、大規模で多様性に富んだ放送映像アーカイブに同一映像断片照合を適用し、これを意味関連グラフととらえて解析することにより、従来の画像解析・画像理解に頼らない、新たな映像意味解析・映像アーカイブ構造化技術の実現を目指す。より具体的には、

- (1)大規模放送映像アーカイブに適用可能な同一映像断片照合の実現、
- (2)同一映像断片照合により抽出された関連性の分類と利用可能な意味的関連性の抽出、
- (3)抽出された関連性によるグラフ構造表現ならびにその解析法、

そしてこれらを通して、

- (4)同一映像断片照合による放送映像アーカイブの意味解析の可能性を示す。

研究成果の概要（英文）：

This project addressed a new technology for video semantic analysis and video archive structuring based on the analysis of semantic relation graph composed of video near-duplicate relation obtained from large-scale and diverse broadcast video archive, thus independent from standard image semantic analysis. In concrete,

- (1) Implementation of near-duplicate detection applicable to large-scale broadcast video archive
- (2) Classification of near-duplicate relation and extraction of semantic relation
- (3) Graph structure representation composed of extracted semantic relation and its analysis
- (4) Revealing potential of semantic analysis of broadcast video archive based of video near-duplicate detection

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	5,600,000	1,680,000	7,280,000
2009 年度	5,100,000	1,530,000	6,630,000
2010 年度	3,900,000	1,170,000	5,070,000
年度			
年度			
総計	14,600,000	4,380,000	18,980,000

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：同一映像断片照合、放送映像アーカイブ、構造化、映像意味解析、映像セマンティックギャップ

1. 研究開始当初の背景

| ブロードバンドネットワークの普及、大容量

記憶装置の普及、デジタル放送の進展等に伴い、大容量のハードディスクレコーダが急速に広まっており、またNHK 川口アーカイブスや第二日本テレビ、YouTube等の新たな映像サービスが始まっている。このような状況を背景として、家庭ならびにインターネット等で個人が視聴可能な映像が爆発的に増えており、この中から適切に必要な映像を選び出し、必要な情報を得るためには、映像の意味解析ならびに検索等の技術の実現が不可欠である。しかしながら、映像の意味解析は本質的に画像認識の問題であり、きわめて難しく、要求に見合うような精度はいまだに達成できていない。その一方で、近年、Near Duplicate Detection(ここでは同一映像断片照合と呼ぶ)と呼ばれる、「同一」と見なしうる映像区間(同じ資料映像、同じシーン、同じ背景、同じ物体、等)を検出する技術が注目を集めている。映像意味解析に比べてはるかに安定した検出性能が期待できる上、検出された映像区間同士は何らかの「同一」のものを映しており、意味的な関連があるとみなすことができ、これをニュース映像アーカイブに適用したときに、関連する話題の検出・追跡、話題の重要性のランキング、新規の話題の検出などに応用できることが示されている。

2. 研究の目的

本研究では、大規模で多様性に富んだ放送映像アーカイブに同一映像断片照合を適用し、これを意味関連グラフととらえて解析することにより、従来の画像解析・画像理解に頼らない、新たな映像意味解析・映像アーカイブ構造化技術の実現を目指す。より具体的には、

- (1)大規模放送映像アーカイブに適用可能な同一映像断片照合の実現、
 - (2)同一映像断片照合により抽出された関連性の分類と利用可能な意味的関連性の抽出、
 - (3)抽出された関連性によるグラフ構造表現ならびにその解析法、
- そしてこれらを通して、
- (4)同一映像断片照合による放送映像アーカイブの意味解析の可能性を示す。

3. 研究の方法

本研究では、主として以下のようなトピックについて検討を行った。

(1) 物体の動きによる同一シーン検出

従来の同一映像断片照合では、基本的には放送映像中で全く同じ資料映像が使われた部分しか検出できず、これらを意味関係とみなすとしても密度が粗すぎるという問題があった。これに対し、我々は、物体の動きに着目し、異なったカメラで撮影されたような、視点が異なる同一シーンの映像の検出を行

う技術について検討を行った。これにより、異なるチャンネル間の意味関係性等も検出できると期待できる。

(2) 高速映像コピー検出技術

同一映像断片(映像コピー)を大規模の放送映像中から大量に収集するためには、高速の映像コピー検出が必要不可欠である。そこで、Glocal と名付けたコンパクトな画像表現と、画像特徴の空間性を考慮に入れたハッシュ技術により、高速に映像コピー検出を行う技術について検討した。

(3) ニューストピックリランキング

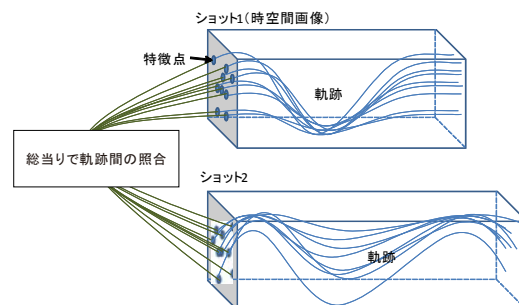
長時間にわたり蓄積されたニュース映像アーカイブ中には、同一映像断片が多く観測できる。そこで、こうした照合結果のうち、ニューストピック間の関連性に相当するものを選び出し、かつこれらにより映像トピックによるグラフ構想が観測できるとみなし、PageRank と同様のアルゴリズムにより、ニューストピックの検索結果のリランキングを行う手法について検討を行った。本成果により、同一映像断片に基づく映像アーカイブの構造化による効果的な応用例が示されている。

(4) コマーシャル映像の検出・同定とマイニング

放送映像アーカイブ中のコマーシャル映像は、特殊な映像コピーとみなすことができる。そこで、映像コピー検出技術を用いることで、コマーシャルデータベースなどの事前知識を用意することなく、自動でコマーシャル映像を検出ならびに同定する技術について検討した。また、その結果に基づくコマーシャル映像マイニング技術により、企業のマーケティング戦略の一端を分析することができることを示している。

4. 研究成果

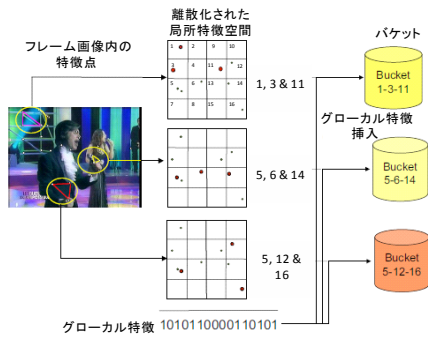
(1) 物体の動きによる同一シーン検出



物体の動きに着目し、撮影の視点が異なるが同一のシーンを撮影しているような映像同士を、一種の「同一映像断片」とみなして検

出す技術を実現した。図に基本的な考え方を示している。各ショットから特徴点を検出し、各特徴点の軌跡を算出し、ショット間の照合の際にはすべての軌跡の組み合わせについて類似性を評価し、一定の割合以上の軌跡が照合可能な場合、元のショットが同一シーンとして検出される。軌跡間の照合において、軌跡の動きが急激に変化している点を検出することにより、我々の以前の研究成果であるフラッシュによる同一シーン検出と同様の手法で二値化系列を抽出し、この系列同士の間での照合によって軌跡の照合を行っている。この手法では、ショット内に一定の動きが必要だが、フラッシュのない映像についても同一シーン検出が可能となり、効果的に同一シーンと同一物体/同一背景との区別ができています。

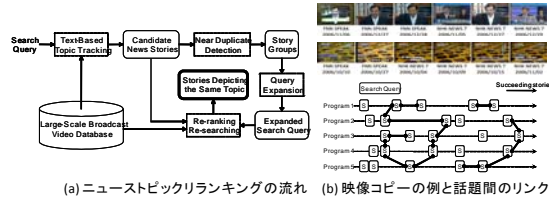
(2) 高速映像コピー検出技術



SIFT などの局所特徴量の利用により、同一資料映像、ならびに強いクロッピング等の映像編集やテロップ重畳等による同一物体/同一背景を含む映像コピー検出を実現した。この手法は照合の高速性に特に注意を払った手法となっている。処理の流れを図に示す。本手法では、キーフレーム画像から 150 点程度の特徴点を抽出し、特徴空間の分割(広く使われている k-means ではなく、特徴空間の軸ごとの分割を利用)し、各セルに特徴点が存在するか否かに基づく二進表現のシグネチャを生成する。これは局所特徴量に基づくものだが、大域的な表現となっているため、ローカルの特性を持ったグローバルな特徴量という意味でグローバル(Glocal)特徴と呼んでいる。次に、照合処理の高速化のため、求めた特徴をハッシュ構造に投入する。ハッシュ関数としては、画像内で相互に一定の近傍内にある 3 特徴点のすべての組と、対応するセル位置による三つ組を利用する。照合処理は、問い合わせの画像・映像からも同様に Glocal 特徴とすべての三つ組を求め、三つ組みに対応するバケットを探索し、それに対応するデータベース中の画像に対し、Glocal 特徴が十分に類似していればコピーとみなす。三つ組を使うことにより、バケットの大部分は空で

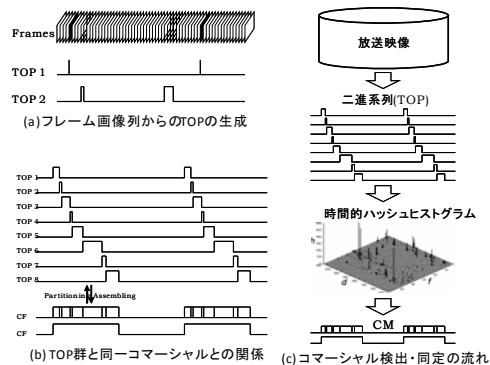
ある。これにより、大部分のキーフレームについては、すべての三つ組に対応するバケットが空となり、高速の照合が可能となる。2009 年の TRECVID 映像コピー検出では、相応の検出精度を実現している手法の中で最高速度を達成した。

(3) ニューストピックリランキング



ニュース映像のトピックの検索システムを考えた場合、キーワード等により与えられる問い合わせに対し合致するトピックを検出するにはテキストに基づく手法を用いるのが有効であると考えられるが、こうしたトピックの中でも特にユーザの意図により合ったものを結果リストの上位に位置づけるようなリランキング技術が重要であると考えられる。そこで、テキストにより検出したトピック群に対し、映像コピー検出技術を適用し、その結果により効果的にリランキングが可能であることを示している。映像コピー検出の結果を Web におけるリンク相当の情報とみなし、グラフ構造を抽出し、PageRank 技術と類似の方法を使ってリランキングを実現し、6 チャンネル 15 か月分という大規模のニュース映像アーカイブ(約 35000 トピック)を対象に実験を行い、映像コピー検出を用いたニューストピックのリランキングの有効性を示した。

(4) コマーシャル映像の検出・同定とマイニング



放送映像中のコマーシャルの検出・同定は、古くから映像コピー検出の重要なアプリケーションの一つであった。これらの技術では、検出すべきコマーシャル映像はあらかじめ与えられ、それによりデータベースが構築

された状態を仮定し、別途提供される放送映像アーカイブからデータベース中のコマーシャル映像を検出することにより目的を実現していた。これに対し、近年、新たな考え方に基づくコマーシャル検出・同定技術が提案されている。最近の手法では、コマーシャル映像があらかじめ提供されることは想定せず、与えられるのは放送映像アーカイブであり、その中に繰り返し現れる映像区間をコマーシャルとみなして自動的に検出する。これにより、コマーシャル映像のデータベースも自動構築できる上、その放送映像中での検出・同定も可能となる。我々はこれを特段に高速化し、知る中で世界最高速の手法を実現した。図に処理の流れを示す。最初に映像アーカイブのすべてのフレーム画像からごく小さいシグネチャを抽出する(ここでは特に32ビットのシグネチャを使用)。シグネチャとしては、同一のコマーシャル映像間の対応するフレーム画像同士など、ほとんど同じとみなしうる画像からは同一のシグネチャが生成される確率が極めて高いような手法であれば任意のシグネチャが使用可能である。そのうち、同一のシグネチャを持つフレーム画像群ごとに、その映像アーカイブ中での時間位置を表す二値化系列(特に大規模の映像アーカイブの場合には極めて長大な系列となる)を抽出する。これを TOP(Temporal Occurrence Pattern)と呼ぶ(図参照)。コマーシャル映像については、出現回数に応じて複数のピークを持つ TOP が生成され、かつ同一のコマーシャルから生成された TOP 同士は、よく似た開始位置並びにほとんど同じピーク間隔を持つため、こうした TOP をまとめることによってコマーシャルの検出と同定が実現できる。コマーシャルと一般の映像コピーとは、長さによって容易に分類可能である。生成された TOP 群に対し、一定程度の時間オフセットを許すハミング距離によりクラスタリングすることでコマーシャルに対応する TOP をまとめることでコマーシャル検出・同定を実現した。さらに、TOP 中のすべてのピーク対について、その時間位置並びに間隔を特徴量とした二次元空間の投票を行い、その結果現れるピークがコマーシャルに対応する TOP 群に対応することを利用し、極めて高速のコマーシャル検出・同定を実現している。シグネチャ生成時間を別にすると、一か月分の放送映像アーカイブ(720時間)からすべてのコマーシャルを検出・同定するの計算時間がわずか 60 分弱であった。その精度も極めて高いことが確認されている。

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) Masaki Takahashi, Mahito Fujii, Masahiro Shibata, Shin'ichi Satoh: "Robust Recognition of Specific Human Motions in Crowded Surveillance Video Sequences" EURASIP J. on Adv. in Sig. Proc., the Special Issue on Video Analysis for Human Behavior Understanding Vol. 2010, Article ID 801252 (2010) (査読あり)
- (2) Masaki Takahashi, Mahito Fujii, Masahiro Shibata, Nobuyuki Yagi, Shin'ichi Satoh: "Automatic Pitch Type Recognition System from Single-View Video Sequences of Baseball Broadcast Videos" the International Journal of Multimedia Data Engineering and Management Vol.1, No.1. 12-36 (2010) (査読あり)
- (3) 武小萌, 瀧本政雄, 佐藤真一, 安達淳: "特徴点軌跡の不均一性パターンに基づいた同一場面映像検出" 電子情報通信学会論文誌 Vol.J92-D, No.8. 1153-1165 (2009) (査読あり)
- (4) 佐野雅規, 八木伸行, 片山紀生, 佐藤真一: "蓄積されたニュース番組からの画像付きクイズ生成手法の提案" 電子情報通信学会論文誌, Vol.J92-D, No.1. 141-152 (2009) (査読あり)

[学会発表] (計 40 件)

- (1) Nobuyuki Morioka and Shin'ichi Satoh, "Compact Correlation Coding for Visual Object Categorization," Intern'l Conf. on Computer Vision (ICCV2011) (2011)
- (2) Thanh Duc Ngo, Duy-Dinh Le, and Shin'ichi Satoh, "Improving Image Categorization by Using Multiple Instance Learning with Spatial Relation," Intern'l Conf. on Image Analysis and Processing (ICIAP 2011) (2011)
- (3) Hiep Van Hoang, Duy-Dinh Le, Shin'ichi Satoh, and Quang Hong Nguyen, "Improving Retake Detection by Adding Motion Feature," Intern'l Conf. on Image Analysis and Processing (ICIAP 2011) (2011)
- (4) Xiaomeng Wu and Shin'ichi Satoh, "Commercial Mining Based on Temporal Recurrence Hashing Algorithm and Bag-Of-Fingerprints Model," Intern'l Conf. on Image Processing (ICIP2011) (2011)

- (5) Thanh Duc Ngo, Duy-Dinh Le, and Shin'ichi Satoh, "Boosting global scene classification accuracy by discriminative region localization," Intern'l Conf. on Image Processing (ICIP2011) (2011)
- (6) Xiao Zhou, Cai-Zhi Zhu, and Shin'ichi Satoh, "Efficient quantization of color sift for image classification," Intern'l Conf. on Image Processing (ICIP2011) (2011)
- (7) Masaki Takahashi, Mahito Fujii, Masahide Naemura, and Shin'ichi Satoh, "Human Action Recognition in Crowded Surveillance Video Sequences by Using Features Taken from Key-Point Trajectories," The 3rd Intern'l Workshop on Machine Learning for Vision-based Motion Analysis, in conjunction with CVPR 2011 (2011)
- (8) Xiaomeng Wu and Shin'ichi Satoh, "Temporal Recurrence Hashing Algorithm for Mining Commercials from Multimedia Streams," Intern'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2011), 2011.
- (9) [特別講演] 佐藤 真一, "映像内容解析ならびに検索技術の最近の動向: TRECVID を中心として," 第5回音声ドキュメント処理ワークショップ, 3/7 (2011)
- (10) Invited Talk, Shin'ichi Satoh, Introduction to Content-based Media Analysis and Search Technology: Technology Overview and Historical Trends from an Academic Point of View, Digital Media Analysis, Search and Management DMASM 2011 Intern'l Workshop (2011)
- (11) [特別講演] 佐藤 真一, 篠田 浩一, "映像解析・検索評価ワークショップ TRECVID2010 の概要," 埼玉大学, PRMU, 2/17-18 (2011)
- (12) Norio KATAYAMA, Hiroshi MO, and Shin'ichi SATOH, "News Shot Cloud: Ranking TV News Shots by Cross TV-Channel Filtering for Efficient Browsing of Large-Scale News Video Archives," Intern'l Conf. on MultiMedia Modeling (MMM2011) (2011)
- (13) Shin'ichi Satoh, "Multimedia Web Analysis Towards Development of Social Analysis Software," The Search Engine Day, Organized by INRIA and Cap Digital (2010)
- (14) Shin'ichi Satoh, Keynote: "Towards Mining Broadcast Video Archive by Latent Link Analysis," Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval (VLSMCMR2010), in conjunction with ACM Multimedia (2010)
- (15) Panel: "Very Large Scale Multimedia Corpus, Mining and Retrieval," Panelist: Edward Chang (Google), Apostol Natsev (IBM), Remi Landais (Exalead), Shin'ichi Satoh (NII), Session Chair: Benoît Huet, VLSMCMR2010, in conjunction with ACM Multimedia(2010)
- (16) Xiaomeng Wu, Narongsak Putpuek, and Shin'ichi Satoh, "Commercial Film Detection and Identification Based on a Dual-Stage Temporal Recurrence Hashing Algorithm," VLSMCMR2010, in conjunction with ACM Multimedia (2010)
- (17) Nobuyuki Morioka and Shin'ichi Satoh, "Learning Directional Local Pairwise Bases with Sparse Coding," Proc. of British Machine Vision Conference (BMVC2010) (2010)
- (18) Tuan Hue Thi, Li Cheng, Jian Zhang, Li Wang, and Shin'ichi Satoh, "Human Action Recognition and Localization in Video Using Structured Learning of Local Space-Time Features," Intern'l Conf. on Advanced Video and Signal-Based Surveillance (AVSS2010) (2010)
- (19) Nobuyuki Morioka and Shin'ichi Satoh, "Building Compact Local Pairwise Codebook with Joint Feature Space Clustering," Proc. of European Conf. on Computer Vision (ECCV2010) (2010)
- (20) Narongsak Putpuek, Nagul Cooharajanone, Shin'ichi Satoh, Chidchanok Lursinsap: "Unified Approach to Detection and Identification of Commercial Films by Temporal Occurrence Pattern" Intern'l Conf. on Pattern Recognition(ICPR2010) (2010)
- (21) Sebastien Poullot, Shin'ichi Satoh: "Detecting screen shot images within large-scale video archive" Intern'l Conf. on Pattern Recognition(ICPR2010) (2010)
- (22) Shin'ichi Satoh, Keynote: "Indexing Video Archives Using Latent Link Analysis," Intern'l Workshop on

- Content-Based Multimedia Indexing (CBMI2010) (2010)
- (23) Panel: "Content-based multimedia search: what about large scale challenges?," Panelist: Shin'ichi Satoh, Allan Hanbury, and Benoit Huet, Moderator: Nozha Boujemaa (INRIA), Intern'l Workshop on Content-Based Multimedia Indexing (CBMI2010) (2010)
- (24) Thao Ngoc Nguyen, Thanh Duc Ngo, Duy-Dinh Le, Shin'ichi Satoh, Bac Hoai Le, and Duc Anh Duong, "An Efficient Method for Face Retrieval from Large Video Datasets," CIVR (2010)
- (25) Xiaomeng Wu, Ichiro Ide, Shin'ichi Satoh: "PageRank with Text Similarity and Video Near Duplicate Constraints for News Story Re-ranking" Special Session on Cross-Media, Cross-Domain and Cross-Source Machine Tagging in Large Scale, Intern'l Conf. on MultiMedia Modeling(MMM2010). (2010)
- (26) Xiaomeng Wu, Ichiro Ide, Shin'ichi Satoh: "News Topic Tracking and Re-ranking with Query Expansion Based on Near-Duplicate Detection" Proc.of Pacific-Rim Conf. on Multimedia(PCM2009) (2009)
- (27) Sebastien Poullot, Michel Crucianu, Shin'ichi Satoh: "Indexing Local Configurations of Features for Scalable Content-Based Video Copy Detection" Intern'l Workshop on Large-Scale Multimedia Retrieval and Mining in, conjunction with ACM Multimedia 2009 (2009)
- (28) Xiaomeng Wu, Ichiro Ide, Shin'ichi Satoh: "Large-Scale News Topic Tracking and Key-scene Ranking with Video Near-Duplicate Constraints" Intern'l Workshop on Large-Scale Multimedia Retrieval and Mining, in conjunction with ACM Multimedia 2009 (2009)
- (29) Duy-Dinh Le, Shin'ichi Satoh: "Unsupervised Face Annotation by Mining the Web" Intern'l Conf. on Data Mining (ICDM2008) (2008)
- (30) Duc Thanh Ngo, Duy-Dinh Le, Shin'ichi Satoh Anh Duc Duong: "Robust Face Track Finding in Video Using Tracked Points" Intern'l Conf. on Signal-Image Technology

Internet-based Systems (SITIS2008), Information Managment and Retrieval Technologies Track (IMRT) (2008)

- (31) Xiaomeng Wu, Masao Takimoto, Shin'ichi Satoh, Jun Adachi: "Scene Duplicate Detection Based on the Pattern of Discontinuities in Feature Point Trajectories" ACM Multimedia (2008)

〔図書〕 (計 2 件)

- (1) Shin'ichi Satoh: "Video Near-Duplicate Detection," in Borko Furht ed., Encyclopedia of Multimedia, 2nd Edition, Springer, 2008.
- (2) Duy-Dinh Le, Xiaomeng Wu, Shin'ichi Satoh: "Face Detection, Tracking, and Recognition for Broadcast Video," in Borko Furht ed. Encyclopedia of Multimedia, 2nd Edition, Springer, 2008.

〔産業財産権〕
なし

〔その他〕
ホームページ等
<http://research.nii.ac.jp/~satoh/>

6. 研究組織

(1) 研究代表者

佐藤 真一 (SATOHI 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 (Duy-Dinh LE)
国立情報学研究所・コンテンツ科学研究系・
助教
研究者番号 : 80450170

武小萌 (Xiaomeng WU)

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