様式 C-19

科学研究費補助金研究成果報告書

平成 22 年 6 月 2 日現在

研究種目： 若手研究（B）
研究期間：2008～2009
課題番号：２０７０００８８
研究課題名（和文）第２世代ビデオマイニング：パターンに基づく映像アーカイブの網羅的検索の実現

研究課題名（英文）Second-generation Video Mining: Extracting Patterns for Exhaustive Retrieval of Events in Video Archive

研究代表者
白浜　公章（KIMIAKI SHIRAHAMA）
神戸大学・大学院経済学研究科・助教
研究者番号：３０４６７６７５

研究成果の概要（和文）：本研究では、データマイニング技術を映像処理に応用して、大量の映像が蓄積された映像アーカイブから、所望のイベントを効率的に検索するための 3 つの手法を開発した。1 つ目は、ラフ集合理論と呼ばれる技術を用いて、所望のイベントに特有の特徴量（色、エッジ、動きなど）の組み合わせをパターンとして抽出する手法を開発した。2 つ目は、イベント中に出現する概念（人、車、建物など）をビデオオントロジーとして体系化し、概念間の関係性に基づいて検索精度を向上させる手法を開発した。3 つ目は、異常な編集パターンが使用されている映像区間を印象的なトピックとして抽出する手法を開発した。

研究成果の概要（英文）：In this research, by applying data mining technique to video processing, we have explored three approaches for efficiently retrieving events of interest in a video archive. In the first one, we use rough set theory to extract combinations of features (e.g. color, edge, motion etc.) specific to events as patterns. In the second approach, we organize various concepts (e.g. Person, Car, Building etc.) into a video ontology. It is used to improve the event retrieval performance. In the final approach, we extract impressive topics in a video by detecting abnormal editing patterns.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	1,900,000	570,000	2,470,000
2009 年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総　計	3,300,000	990,000	4,290,000

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（1004）A
キーワード：映像検索、ビデオマイニング、ラフ集合理論、バースト検出、ビデオオントロジー、TRECVID

１．研究開始当初の背景

近年，YouTube やニコニコ動画といった動画共有サイトに代表されるように，インターネットを介して，大量の映像が蓄積された“映像アーカイブ”にアクセス可能になってきた．映像アーカイブでは，映像の本数もさ

ることながら，時間を伴うメディアであるため，視聴に時間を要するという問題がある．そのため，映像アーカイブから，例えば「人が街中を走っている」，「山中の城が映っている」といった，ユーザが所望する部分映像（以下，“イベント”）を効率的に検索可能なシステムの技術開発に対する期待が非常に高まっている．

本研究では，研究代表者らが，2003 年から世界に先駆けて研究を行っている“ビデオマイニング”というアプローチを映像アーカイブ検索に応用する．ビデオマイニングとは，データマイニング技術を用いて，映像からイベントを検索するために有用なパターンを発見する技術である．特に，従来のビデオマイニングでは，小規模な映像データ中の限られた種類のイベントを検索するパターンしか抽出できなかったのに対して，本研究では大規模な映像データを対象として，多種多様なイベントを検索するためのパターンを抽出可能な“第２世代ビデオマイニング”技術を開発する．

２．研究の目的

映像アーカイブに対しては，多種多様なイベントに関する検索要求がある．しかしながら，映像アーカイブの規模を考えると，考えうる全てのイベントをあらかじめ索引付けしておくことは不可能である．また，YouTube やニコニコ動画などのように，映像中の代表的な意味内容だけを索引（タグ）付けしたのでは，明らかに不十分である．そこで，本研究では，ユーザが所望のイベントに関するサンプル映像をクエリーとして与え，サンプル映像中の色，エッジ，動きといった特徴量を自動解析して，検索モデルを動的に生成するアプローチをとる．これによって，サンプル映像さえあれば，映像アーカイブ中の任意のイベントを検索可能になる．

ただし，色やエッジといった物理的な特徴量だけでは，高精度なイベント検索を実現することは困難である．そこで，人手により用意された意味的な情報を背景知識として利用した検索手法を開発する．具体的には，イベントに出現する概念間の関係性を考慮して，例えば「船が映っているイベント」を検索するために，「船」の他に「海」も映っているかどうか検証すれば，検索精度の向上が期待できる．そこで，イベント中に含まれる概念，概念の性質，概念間の関係性を“ビデオントロジー”として体系化する．

映像は時間を伴うメディアであり，長い映像であれば，1 本を視聴するのに，数時間要する．そのため，ユーザが映像の意味内容を短時間で把握するためのブラウジング技術には非常に大きな意義がある．そこで，本研究では，映像を意味的にまとまりのある映像区間に分割し，印象的な意味内容が表現された“トピック”を抽出する手法を開発する．

３．研究の方法

以下では，「(1)サンプル映像からのイベント検索モデルの導出」，「(2)ビデオントロジーの構築」，「(3)トピック抽出」のそれぞれに関する研究方法を説明する．

(1) サンプル映像からのイベント検索モデルの導出

まず，本研究では，ユーザから２種類のサンプル映像が与えられることを前提としている．１種類目は所望のイベントが表現された映像（“正例”）であり，２種類目は所望のイベントが表現されていない映像（“負例”）である．そして，正例と負例を比較して，正例に特有の特徴量を抽出し，検索モデルを生成する．

検索モデル生成過程において，「撮影・演出技法によって，同一イベントの映像でも特徴量が大きく異なってくる」という多様性に着目する．例えば，図１の「街中を車が走っている」イベントは，*shot 1* のような近距離から撮影された映像では，画面全体の色変化（動き）が大きい．また，*shot 2* のような遠距離でかつ郊外で撮影された映像では，画面の一部から大きな動き，画面上部の空に対応した青色といった特徴量が検出される．一方，*shot 3* のような都市部で撮影された映像では，画面上部のビルに対応して，直線状のエッジが多数検出される．このような考察から，同一イベントのショットは，特徴空間中の一カ所には固まって分布しているのではなく，部分集合に分かれて分布していると仮定できる．

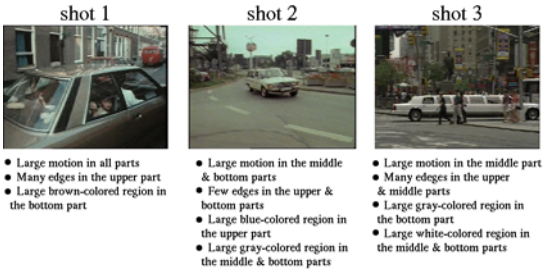

図 1：同一イベントにおける特徴量の多様性

上記のような特徴量の多様性を考慮して本研究では，“ラフ集合理論”を用いて，イベント検索モデルを抽出する．ラフ集合理論とは，明確に定義できないクラスの事例を，正確に定義可能な部分集合ごとに部分的に

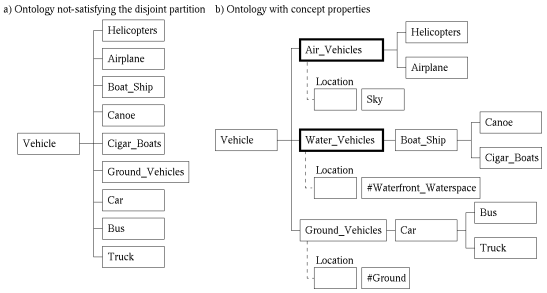

図 2：排他性に基づくビデオオントロジーの構築

定義する教師つき学習手法である． すなわち，正例のある部分集合を検索するためには検索モデル A，別の部分集合を検索するためには検索モデル B というように，同一イベントに対して複数の検索モデルが抽出される．最終的に，抽出された検索モデルを用いて，多角的にイベントを検索可能になる．

(2) ビデオオントロジーの構築

まず，映像処理の分野で標準的に使用されている 374 個の概念を体系化する．ここで，一般的なオントロジー構築のデザインパターンとして知られる“排他性”を考慮する．排他性とは，ある概念のインスタンスが，複数の下位概念のインスタンスになってはならないという制約である．例えば，図 2(a)では，*Vehicle* という概念の下位概念として，*Car*，*Bus*，*Truck* などが列挙されているが，*Truck* のイスタンスは，*Car* のインスタンスでもあるため排他性に反している．そこで，図 2(b)のように，*Car* の下位概念として *Bus*，*Truck* を配置するというように，概念間の上位・下位関係を考慮しながら，ビデオオントロジーを構築する．

また，図 2(b)で，*Air_Vechile* の移動空間（*Location*）は *Sky* であるというように，各概念の性質を定義する．特に，*Air_Vehicle* に対して *Sky* というような意味的に明らかな性質だけでなく，*Building* に対して *Sky* というような画面上に頻繁に共起して出現する概念も性質として定義する．さらに，標準的な 374 個の概念だけでは，意味のある階層関係を定義できない場合は，新規概念を追加する．例えば，図 2(c)では，*Air_Vehicle* や *Water_Vehicle* を新規概念として追加している．

最後に，以下のようにして，構築したビデオオントロジーをイベント検索に使用する．まず，ショットごとに，各概念の出現確率を表す認識結果があるとする．今，ユーザから正例と負例が与えられたとして，ラフ集合理論を用いて，正例に出現して，かつ負例に出現していない概念の組み合わせを検索モデルとして抽出する．この過程において，所望のイベントに関連する概念を絞り込むためにビデオオントロジーを使用する．例えば，「車が街中を走っている」イベントを検索するためには，*Parade* や *Weapons* といった概念は明らかに不適切である．しかしながら，正例と負例，概念認識のエラーによっては，検索モデルに明らかに不適切な概念が含まれてしまい，精度低下の原因になる．このように本研究では，ビデオオントロジーをイベント検索の際の特徴選択のために使用する．最終的に，検索モデルに当てはまった映像が検索結果として，ユーザに提示される．

(3) トピック抽出

トピック抽出に当たっては，まず，映像編集技術に着目する．図 3 に示されているように，映像は単一のカメラから撮影されたショットと呼ばれる断片映像をつなぎ合わせて制作されている．特に，プロの映像編集者は，登場人物に起こった緊迫した出来事を表現するために，その人物のアクションを時間長の非常に短いショットに細かく区切るという編集技法を使用する（図 3 の一番右のトピック）．一方，叙情的な出来事を表現するためには，時間長の非常に長いショットを用いて，登場人物のアクションをじっくりと映すという編集技法が使用されている．このような編集技法に基づいて，映像中で，登場人物の出現時間が異常に短くなる，もしくは長くなるという異常パターン（“バースト”）を検出して，視聴者にインパクトを与える印象的なトピックを抽出する．

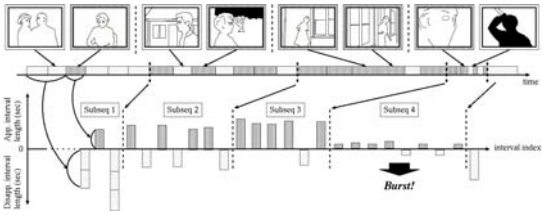

図 3：バースト検出に基づくトピック抽出の概要

トピック抽出手法としては，まず，映像を，図 3 の真ん中に示されているような登場人物の出現時間と非出現時間を表すシークエンスに変換する．そして，確率モデルに基づく時系列セグメンテーション手法を用いて，登場人物の出現パターンが類似したサブシークエンスに分割する．最後に，サブシークエンスごとに，登場人物の出現パターンのバースト性を検証して，バーストが発生していると判定されたサブシークエンスをトピックとして抽出する．

４．研究成果

以下では，「(1)サンプル映像からのイベント検索モデルの導出」，「(2)ビデオオントロジーの構築」，「(3)トピック抽出」のそれぞれに関する研究成果について報告する．

(1) サンプル映像からのイベント検索モデルの導出

まず，実験映像として，米国 NIST 後援の研究プロジェクト TRECVID（2008 年時点）から提供された 439 本の映像データ（合計 71,872 ショット）を使用した．そして，表 1 に示されているように，「車が街を走る」，「街中でのインタービュー」，「ドアを開ける」という 3 種類のイベントを検索した．2 行目と 3 行目に検索に使用した正例，負例の数を示す．検索性能の評価方法としては，4 から 7 行目に示されているように，検索時の評価値に基づいてショットをランク付けし，上位 100，300，500，1000 位以内に何個の正解が含まれているか算出した．

表 1：イベント検索の性能

検索数	車が街を走る	街中でのインタビュー	ドアを開ける
正例数	9	11	9
負例数	14	16	16
p@100	0.340	0.100	0.090
p@300	0.216	0.087	0.070
p@500	0.168	0.078	0.068
p@1000	0.139	0.068	0.063

表 1 の検索性能は，世界の各研究機関で開発された映像検索手法と比べると，良好であるとは言えないが，図 4 から分かるように，ラフ集合理論を用いて，同一のイベントでも，ショットサイズや撮影状況が大きく異なる多様なショットを検索できることを確認した．また，現在は，本研究課題実施時の手法に更に改良を加え，世界の水準以上の検索精度が得られるようになっている．最後に，本検索手法に関しては，4 つの国際会議（学会発表の[2]，[6](査読有)，[3]，[4](査読無)），2 つの国内研究会で研究発表(学会発表の[5]，[7])を行った．

図 4：検索されたショットの例

(2) ビデオオントロジーの構築

実験内容としては，TRECVID（2009 年時点）から提供された 639 本の映像データ（合計 133,256 ショット）を対象として，「高い建物が映っている」，「電話をしている」，「炎が燃えている」などの合計 8 種類のイベントを検索した．そして，374 個の概念全てを用いて検索した場合と，ビデオオントロジーを用いてイベントに関連する概念を絞り込んだ場合での検索性能を比較した（表 2 の 2，3 行目）．ここで，検索された 1000 ショット中に何個の正解が含まれているか数えて，検索性能を評価している．また，概念の認識結果に加えて，視覚的な特徴量を併用した場合の検索性能についても検証した（表 2 の 4 行目以下）．

表 2：ビデオオントロジーを用いたイベント検索性能の評価

	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6	Event 7	Event 8
All concepts	37	204	104	9	97	23	25	80
Selected concepts	**43**	**280**	101	6	**236**	**33**	**39**	**126**
Color	31	187	111	5	197	30	13	54
Edge	50	124	108	5	162	26	46	63
SIFT	26	157	106	9	441	15	8	61
Moving region	40	280	97	7	194	21	42	91
Face appearance	49	275	100	6	170	42	40	101
Gabor texture	40	172	125	14	336	26	34	116
Color moment	19	227	119	5	31	12	10	67
Camera work	27	239	104	15	73	9	15	101

表 2 の 2，3 行目から，8 個中 6 個のイベントで，ビデオオントロジーを用いて検索した方が高い精度が得られている．特に，「Event5: 印刷，タイプ，手書きの文字が映っている」イベントでは，ビデオオントロジーを用いることで精度が 2 倍以上向上している．この結果から，構築したビデオオントロジーの有効性が示せたと言える．また，表 2 の 4 行目以下の太線の四角で示されているように，8 個中 5 個のイベントで，概念の認識結果と視覚的な特徴量を併用することで検索精度が向上している．特に，Event 5 では，2 倍近い精度向上があった．

上記の結果から，人手により与えられた知識を映像検索に利用することが非常に有用であることを示している．今後の拡張点として，ショットサイズやカメラワークなど映像特有の情報をビデオオントロジーに導入すること，YouTube などの動画共有サイトで既に付与されているタグ情報から大規模なビデオントロジーを自動構築する手法を開発することなどが挙げられる．最後に，上記のビデオオントロジーに関しては，研究実施期間中に研究発表を行うことができなかった．

(3) トピック抽出

バースト検出に基づくトピック抽出に関しては，4 本の商用映画を実験映像とした．そして，各映像の主人公の出現時間と非出現時間を表すシークエンスを作成し，提案手法によりトピックを抽出した．図 5 に抽出され

たトピックの例を示す．ここで，上記のシークエンスを棒グラフ形式で表現している．具体的には，正方向の棒が登場人物の出現時間，負方向の棒が非出現時間を表している．また，各サブシークエンスにおける登場人物の出現パターンのバースト性の評価値を，太線のパルス波の形式で表している．すなわち，この評価値が高いサブシークエンスが，バーストが発生しているトピックとして抽出される．

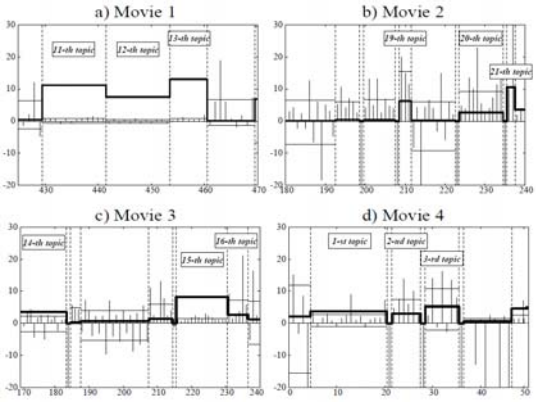

図 5：バースト検出に基づくトピック抽出結果

まず，図5から，提案手法により，登場人物の出現時間と非出現時間が類似したサブシークエンスに適切に分割できていることが分かる．ここで，ユーザにとって意味のあるトピックを抽出するためには，各サブシークエンスに意味的なまとまりがあることが望ましい．この点に関して，分割されたサブシークエンスの 77%が意味的なまとまりがあるシーンに対応していることを確認した．すなわち，登場人物の出現時間と非出現時間という情報だけでも，映像を意味的にまとまりのあるシーンに高精度に分割できる．最終的に，抽出されたトピックを検証した結果，殺人，戦闘，恋愛，ダンスなど，映像の中でも特に印象的なトピックが抽出されていることが分かった．

現在は，本手法を自動化するために，登場人物の自動認識手法について検討している．本手法が自動化されれば，トピックに基づく映像ブラウジングだけでなく，人物の出現パターンによって，プロが作成した映像と素人が作成した映像を判別できる可能性があり，著作権侵害コンテンツの検出にも利用可能である．最後に，本トピック抽出手法に関しては，1件の雑誌論文（雑誌論文の[1]）と1件の国際会議（学会発表の[1]）で，研究成果を発表した．ここで，前者の雑誌論文は，後者の国際会議で発表した論文が Selected Papter として選ばれたため執筆したものである．

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 1 件）

[1] Kimiaki Shirahama and Kuniaki Uehara, A Novel Topic Extraction Method based on Bursts in Video Streams, 査読無, International Journal of International Journal of Hybrid Information Technology (IJHIT), Vol. 2, No. 3, 2008, pp. 21 - 32. （学会発表[1]からのSelected paper）

〔学会発表〕（計 7 件）

[1] Kimiaki Shirahama, Chieri Sugihara and Kuniaki Uehara, Query-based Video Event Definition Using Rough Set Thoery, The First ACM International Workshop on Events in Multimedia (EiMM 2009), October 23, 2009, Beijing Hotel, Beijing, China. （査読有）

[2] 杉原ちえり，白浜公章，上原邦昭，ラフ集合理論を用いたクエリー映像からのイベント検索モデルの導出，平成21年度　情報処理学会関西支部　支部大会，2009年9月29日，神戸大学．（査読無）

[3] 水井章人，白浜公章，上原邦昭：”多重対応分析を利用した特徴量選択による映像検索精度の改善”，2009 年電子情報通信学会総合大会，D-12-37，3 月 18 日,2009.（査読無）

[4] Akihito Mizui, Kimiaki Shirahama and Kuniaki Uehara, TRECVID 2008 NOTEBOOK PAPER: Interactive Search Using MultipleQueries and Rough Set Theory, TREC Video Retrieval Evaluation (TRECVID) 2008 Workshop, November 17, 2008, National Institute of Standards and Technology (NIST), Meryland, US. （査読無）

[5] Kimiaki Shirahama, Akihito Mizui, and Kuniaki Uehara, Characteristics of Textual Information in Video Data from the Perspective of Natural Language Processing, NSF Sponsored Symposium on Semantic Knowledge Discovery, Organization and Use, November 15, 2008, New York University, New York, US. （査読無）

[6] Kimiaki Shirahama and Kuniaki Uehara, Query by Shots: Retrieving Meaningful Events Using Multiple Queries and Rough Set Theory, The 9-th International

Workshop on Multimedia Data Ming (MDM/KDD 2008), August 24, 2008, Loews Lake Las Vegas Resort, Nevada, US. (査読有)

[7] Kimiaki Shirahama and Kuniaki Uehara, A Novel Topic Extraction Method Based on Bursts in Video Streams, The 2-nd International Conference on Mulitimedia and Ubiquitous Engineering (MUE 2008), April 24, 2008, Hanwha Resort Haeundae, Busan, Korea. (査読有)

〔図書〕(計0件)

〔その他〕
ホームページ等
[1]http://www.ai.cs.scitec.kobe-u.ac.jp/research_html/trecvid/video_retrieval.html

[2]http://www-nlpir.nist.gov/projects/tvpubs/tv8.papers/cs24_kobe.pdf

[3]http://www-nlpir.nist.gov/projects/tvpubs/tv9.papers/cs24_kobe.pdf

[4]http://www-nlpir.nist.gov/projects/tvpubs/tv9.slides/kobe.slides.pdf

６．研究組織
(1)研究代表者
白浜　公章 (SHIRAHAMA KIMIAKI)
神戸大学・大学院経済学研究科・助教
研究者番号：３０４６７６７５