論文 / 著書情報
Article / Book Information

論題(和文)	映像検索技術の新たな潮流
Title(English)	Emerging Trends in Video Search Technology
著者(和文)	篠田浩一
Authors(English)	Koichi Shinoda
出典(和文)	電子情報通信学会誌, Vol. 95, No. 10, pp. 932-938
Citation(English)	, Vol. 95, No. 10, pp. 932-938
発行日 / Pub. date	2012, 10
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 2012 Institute of Electronics, Information and Communication Engineers.

解説

映像検索技術の新たな潮流

Emerging Trends in Video Search Technology

篠田浩一

Abstract

インターネット上に映像があふれている現在，その高性能な検索技術が強く望まれており，多くの研究機関がその開発に取り組んでいる．その中で，米国で毎年開催されている映像検索・評価ワークショップ TRECVID は，世界各国から多くの有力研究機関が参加する競争型国際ワークショップであり，最先端技術のショーケースとなっている．ここでは，その代表的な二つのタスク，オブジェクトやシーンなどの高次特徴を検索する意味インデクシングと，それら高次特徴の組合せとして表現されるイベントを検索するマルチメディアイベント検出について，最新の動向を解説する．

キーワード：映像検索，意味インデクシング，マルチメディアイベント検出

1. はじめに

映像の検索では，文書検索に比べると，自然言語で与えられる検索クエリと検索対象の映像コンテンツの対応がより不明確であり，いわゆる「セマンティックギャップ」の問題がより深刻である．その解決のために，パターン認識技術を活用した映像コンテンツ検索の研究が精力的に行われている．例えば，ACM Multimedia は，映像分野で最も権威ある国際会議でありアート（芸術），符号化，データベース，コンテンツ解析など幅広い分野での発表があるが，2011 年の同会議では映像コンテンツ検索に関する発表が全体の 3 分の 1 以上を占めた．

今までの研究の多くは，映画，ドキュメンタリー，スポーツ番組など，ある程度整備されたデータベースを用い，サッカー中継におけるゴールシーンの検出など，ある限定された用途を対象としてきた．一般的な問題への応用はまだこれからである．特に，インターネット上に映像があふれている現在，その高性能な検索技術が望まれている．これらの映像の多くは，素人により作成されたものであり，その品質は低い．また，多くの場合，検索に必要なテキスト情報（メタデータ）が十分に付与されておらず，付与されていたとしても必ずしも正確でない．更に，検索の対象となる事物は極めて多岐にわたる．このような状況下における映像検索は極めてチャレンジングなタスクであるといえよう．

インターネット映像の検索の研究では，先に述べたように，映像が多様であり検索対象の種類が多いため，個々の技術の性能を比較することが難しい．この問題意識の下，2001 年から米国国立標準技術局（NIST : National Institute of Standards and Technology）の主催で映像検索・評価ワークショップ TRECVID が毎年開催されている[1]．このワークショップは競争型の国際ワークショップであり，映像検索に関する幾つかの共通のタスクが用意され，それに対する性能を競う．毎年，世界各国から多くの有力研究機関が参加している．そこでは，様々な映像検索技術の性能を容易に比較することができ，その最先端技術のショーケースとなっている．

例えば，意味インデクシング（SIN : Semantic Indexing）タスクは，「机」，「夜景」，「歌う」など映像における「単語」をクエリとし，それを含む「映像ショット」を検索するタスクである．2002 年から 10 年以上続いており，参加チーム数も最も多い基盤的かつ中心的なタスクである．また，2010 年から開始されたマルチメディアイベント検出（MED : Multimedia Event Detection）タスクは，「結婚式でケーキを切っている」など複数の「単語」から構成される「文」としての「イベント」を「映像クリップ」から検出するタスクであり，意味インデクシングからの発展を狙う野心的なタスクである．

本稿では，この二つのタスク，意味インデクシングと

篠田浩一　正員：シニア会員　東京工業大学大学院情報理工学研究科計算工学専攻
E-mail　shinoda@cs.titech.ac.jp
Koichi SHINODA, Senior Member (School of Information Science and Technology, Tokyo Institute of Technology, Tokyo, 152-8552 Japan).
電子情報通信学会誌　Vol.95 No.10 pp.932-938 2012 年 10 月

表1 TRECVIDのタスク

Task	タスク	年
Shot boundary detection (SBD)	ショット境界検出	2001～2007
Know item search (KIS)	検索システム（インタフェースも含む）	2001～
Semantic indexing (SIN)	意味インデクシング	2002～
Story segmentation	ストーリー分割	2003～2004
Low-level feature extraction	カメラの動き検出	2005
Rushes summarization	ビデオ要約	2006～2008
Surveillance event detection (SED)	監視カメラからのイベント検出	2008～
Content-based copy detection (CCD)	コピー検出	2008～
Instance search (INS)	特定対象の認識	2010～
Multimedia Event detection (MED)	一般的なイベントの検出	2010～

マルチメディアイベント検出について，その概要と各参加チームの取組みを紹介する．その上で，そこでまさに今現在起きている映像検索技術における革新と，予想される今後の展開について解説する．なお，TRECVIDの詳細については，ホームページ[1]に網羅的な資料があり，本会誌でも過去に解説[2]があるのでそちらを参照されたい．

2. 映像検索・評価ワークショップTRECVID

TRECVID[1),3)]は，NISTの主催で毎年開催されている，映像コンテンツ解析・検索技術の高度化を目的とした，競争型国際プロジェクトである．世界中から研究グループの参加を募り，参加者間で数百時間規模の大規模映像アーカイブを共有すると同時に全員で同じタスクに挑戦し，その結果を比較評価することにより研究水準の向上を図っている．2001年にTREC（テキスト検索の競争型国際プロジェクト）のVideo部門として開始され，2003年からTRECVIDとして独立して開催されている．毎年春にタスクの詳細が決まり，参加チームは夏頃にその結果を提出し，10月頃に各参加チームの成績が通知される．11～12月に開催されるクローズドなワークショップで，参加チームが集まり，その年の成果及び来年度の計画について議論をする．2011年は66チームが結果の提出を行った．日本からも12チームが参加した．

過去から現在に至るまでのタスク名とその内容について表1に簡単にまとめる．例えばショット境界検出（SBD）は映像検索の入門的なタスクとしての位置付けで，2001年の開始時に始まったが，実用の水準に達し使命を終えたと判断され，2007年をもって終結した．また，近年は，監視映像からのイベント検出（SED）やコピー検出（CCD）など，新たなタスクが追加されている．

なお，ヨーロッパでもEU FP7の映像検索・評価のプロジェクトMediaEval Benchmarking Initiative for Multimedia Evaluation[4]が毎年開催されている．こちらはどちらかというと音声関連の研究者が中心で，また，TRECVIDと比べると実用面よりも学術面を重視したタスクが設定されている．残念ながらTRECVIDのコミュニティとの交流はさほど盛んではない．

3. 意味インデクシング

3.1 タスクの概要

意味インデクシング（SIN）タスクは映像のショット（カメラの切り替わりから次の切り替わりまでの間）から「意味」のある物体・シーンを検出することを目的としている．2002年から存続し，最も多くのグループが参加する中核的なタスクとなっている．物体の例としては，Flower，Bus，シーンの例としては，Explosion_Fire（火が燃えている），Car_Racing（カーレース）がある．TRECVIDでは，映像における動画像や音声の信号から抽出される低レベルの特徴を総称して「低次特徴」と呼び，これらの「意味」を持つ物体・シーンなどは「高次特徴」（HLF：High Level Feature）と呼ぶ．ここでもこれらの用語を用いることにする．

このタスクは，あらゆる高次特徴に対して有効な，汎用の高次特徴の検出手法を提供することを目的としている．その意味で画像認識分野で現在盛んに研究されている一般物体検出を映像まで拡張したタスクである．ま

用語解説

Bag of Visual Words (BoW)　量子化した特徴量の頻度ヒストグラムを識別器の入力とする方法の総称．画像中の位置情報を用いない．

Scale Invariant Feature Transform (SIFT)　画像の特徴点において，回転・拡大縮小・照明変動に対し頑健な特徴量を記述するアルゴリズム．

サポートベクトルマシン (SVM)　教師あり学習を用いる識別手法の一つ．パターン認識・回帰分析によく用いられる．特に学習データが少ないときに有効．

た，映像を文とみなしたとき，それを構成する単語を検出する試みと捉えることもできる．すなわち，音声ドキュメントの検索における Spoken Term Detection（STD）と同様の位置付けである．

2010 年から Internet Archive からクリエイティブコモンズライセンスにより提供されている，自由に利用可能なインターネット上の映像からなるデータベース Internet Archive Creative Commons（IACC）を使用している．年々その規模は増加している．2011 年は開発用 400 時間，評価用 200 時間の映像データが用意された．ショット数は開発用が 264,673 個，評価用が 137,327 個である．また，検出対象の高次特徴としては，汎用性（検索における有用性）が高く，出現頻度がある程度大きいものが選ばれる．2011 年は 346 種類であった．開発データにはこれらの高次特徴のラベルが付与される．映像中の空間的・時間的位置については問わず，1 フレーム以上に 1 回でも出現していたら正例となる．一つのショットに複数のラベルが付くこともある．

参加チームは，評価データが配布されたら，346 種類の各々の高次特徴について，評価データの全てのショットに対して検出を試み，検出されたショットを確度の高い順に並べ，順序付きで上位 2,000 位のショット ID を提出する．1 チーム当り計四つの run（検出結果）を提出でき，1 チーム内でも複数の手法を評価することができる．

評価基準としては平均適合率（AP：Average Precision）を用いる．AP は，適合率と再現率を 2 軸とする平面上でしきい値を変化させて得られる曲線（Precision-Recall カーブ）の内側の面積に相当する値で，大きいほどよい．全ての高次特徴の検出の評価には各高次特徴の AP の算術平均である Mean AP が用いられる．Mean AP では高次特徴間の出現頻度の違いは考慮されない．

評価データには高次特徴のラベル付けがされていないため，主催者側は目視で評価を行う必要がある．しかし，全てのショットについて評価するのは工数上困難である．そこで，提出締切後，346 種類の高次特徴の中から 50 種類が選択され，それらが評価の対象となる．選択された 50 種類の高次特徴については図 1 の横軸を参照されたい．また，評価基準として，提出ショットからサンプリングを行い判定した inferred AP（infAP）を用いる．

2011 年は，参加 56 チーム中，28 チームが結果を提出した．各チームから提出された run ごとの結果を図 2 に示す．最高で Mean infAP が 17.3% となった．この値は一見低いように感じるが，上位には明らかな正解が多く含まれている．例えば，ある高次特徴を含む映像ショットが幾つか欲しい，という用途のためには十分で

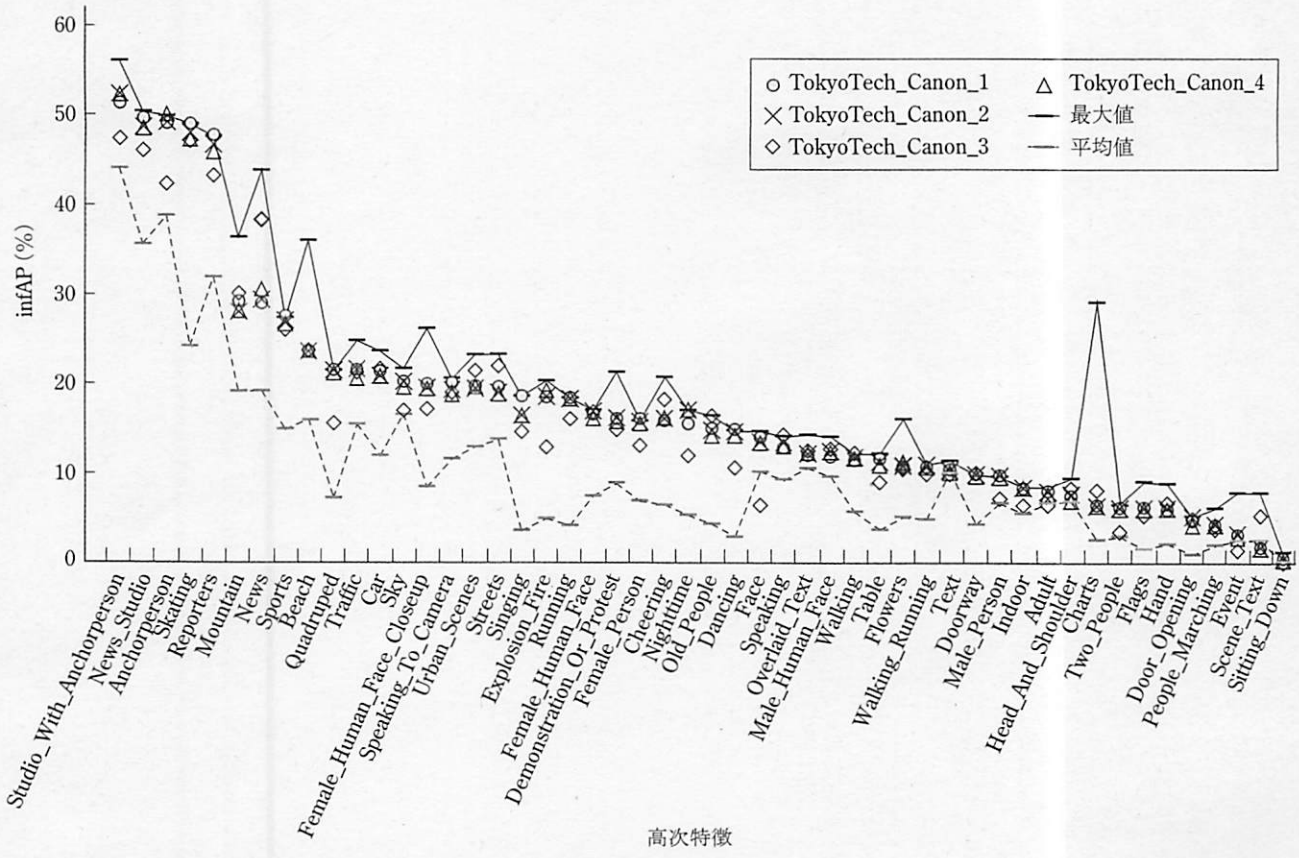

図 1　2011 年意味インデクシングタスクにおける高次特徴ごとの infAP　　TokyoTech_Canon[5],[6] による四つの run の値も併せて示す．

ある（図3）．各高次特徴ごとの結果を図1に示す．高次特徴ごとに性能が大きくばらついていることが分かる．これは，難易度やデータ量の違いによるものである．

3.2 検出フレームワーク

開始から10年が経過し，その間，意味インデクシングに用いられる手法も変遷してきた．基本のフレームワークは静止画像における一般物体認識手法とほぼ同じである．そこでは，各々のショットのキーフレームから抽出した局所特徴のヒストグラムであるBag of Visual Words（BoW）(用語)を入力とし，識別にone-versus-allの識別器を用いる．そして，現在のトレンドは，マルチモーダル，マルチフレーム，マルチカーネル（3M）のキーワードと確率的アプローチである．

3.2.1 基本：一般物体認識手法

ビデオフレームごとの特徴量としては，静止画像の一般物体認識で有効な特徴量がそのまま使われる例が多い．これらの特徴量は画像全体の特徴を表す大局特徴と画像の一部（特徴点）の特徴を表す局所特徴の2種類に大別できる．大局特徴の例としては色ヒストグラムがある．局所特徴の抽出では，まず画像中の特徴点を検出し，その特徴点の周囲のピクセル値を用いて特徴量を抽出する．特徴点検出にはHarris検出器やHessian検出器など，変化の激しい点を抽出する検出器が用いられ，特徴量としては拡大縮小や回転などの操作に対し不変な特徴量が用いられる．SIFT（Scale Invariant Feature Transform）(用語)がその代表的なものであり，その派生としてのColor-SIFT，Opponent SIFT，それ以外にHOG（Histogram of Oriented Gradients），SURF（Speeded Up Robust Feature）などがよく用いられる．また，画像中の位置情報を表現する空間ピラミッド（Spatial Pyramid）もしばしば利用される．

局所特徴は，フレームごとにその数が異なるため，入力特徴量の次元数が決まっている通常の識別器には用いることができない．そこで，BoWが用いられる．そこでは，まず，学習データ全体から抽出された特徴量の集合に対し，ベクトル量子化を行って，コードブックを作成する．ベクトル量子化にはK平均アルゴリズムがしばしば用いられる．そして，入力画像ごとにコードヒストグラムを作成し，それを識別器への入力とする．識別器としてはサポートベクトルマシン(用語)が一般に用いられている．従来は，大局特徴と局所特徴を共に用いることが多かったが，現在は，局所特徴のみが用いられることも多い．これは，局所特徴は一般に大局特徴に比べ情報量が多く，大局特徴に含まれる情報も含有しているためと考えられる．

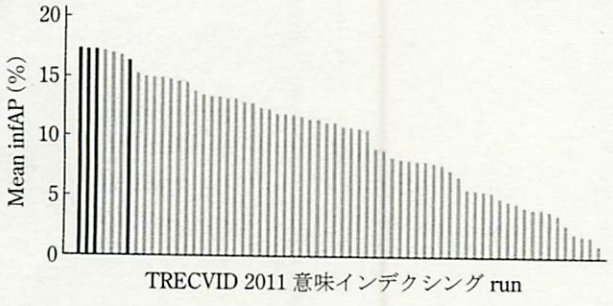

図2 2011年意味インデクシングタスクにおける各チームのMean infAP　濃い線で示されているのはTokyoTech_Canon[5],[6]による四つのrunの値．

3.2.2 マルチモーダル

映像には，画像情報以外に音響情報も含まれている．音声や音楽に関係した高次特徴は，その検出に音響情報

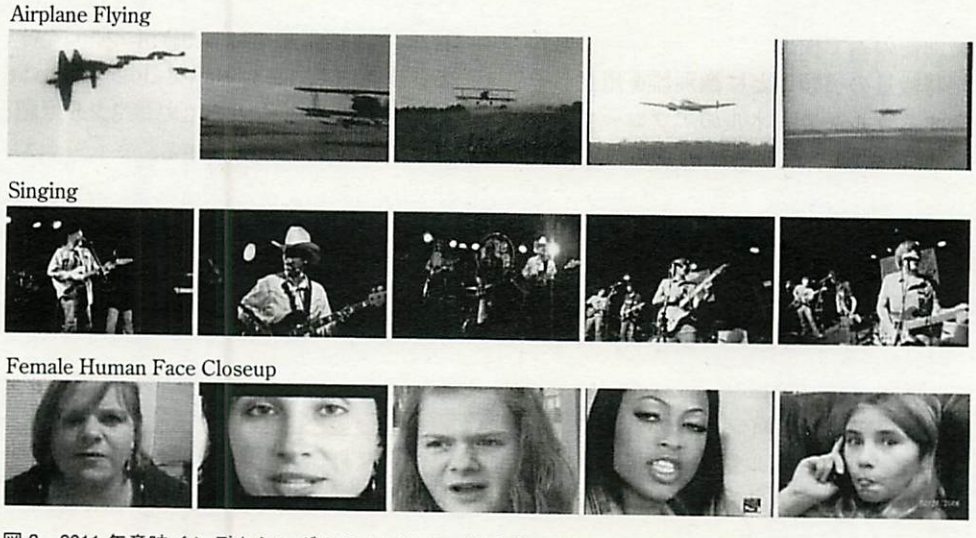

図3 2011年意味インデクシングタスクにおける検出結果の例　3種類の高次特徴について，TokyoTech_Canon[5],[6]による5位までの検出結果を示す．

が重要な役割を果たす．音響特徴量として，音声認識でしばしば用いられるMFCC（Mel-Frequency Cepstral Coefficient）が，この目的にも有効であることが分かっている．MFCCは，音声波形に対する短時間フーリエ変換により得られたパワースペクトルの対数をとり，その結果に対しコサイン変換を行ったものであり，スペクトルの概形を表現する．また，音声認識の場合と同様にMFCCの時間方向差分もしばしば用いられる．

更に，画像・音響情報に含まれる言語（文字）情報を特徴量として使おうという試みも幾つか行われている．例えばOCR，音声認識などである．これらは現在までのところ目立った効果は報告されていない．これは，インターネット映像は極めて多様で雑音が多く，それら自体の性能が低すぎるためである．

3.2.3 マルチフレーム

以前は計算量の制限から，あらかじめ与えられているショット中のキーフレームの画像1枚を学習に用いることが多かった．しかし，当然ながら，映像を構成する複数の画像を学習データとして用いた方が，見えの違いなどの変動に対しより頑健な識別器を構築することができる．計算資源の増大により，多くのフレームを処理することが可能となってきている．計算量との兼ね合いにより，毎フレームごと，1フレームおき，2秒に1フレームなど様々な間隔が用いられる．

3.2.4 マルチカーネル

マルチモーダルのアプローチは，必然的に入力特徴量次元の増大を伴う．また，近年，局所特徴としてDense特徴（Dense SIFT, Dense HOGなど）が，従来の局所特徴とともに用いられることが増えている．これらは特徴点検出は行わずに，あらかじめ定められた画像中の格子点から特徴を抽出するものである．このように入力特徴の次元数が増加すると，主に計算リソースの制限から，単一の識別器のみで検出処理を行うことが難しくなる．そこで，特徴量の種類ごとに識別器を用意しその出力結果を統合するマルチカーネルのアプローチが一般に用いられる．統合の際の識別器間の重みは，開発セットを用いた交差検定（クロスバリデーション）により定められることが多い．

3.2.5 確率的アプローチ

BoWに基づく枠組みでは，特徴量が高次元になるほど大きなコードブックが必要となる．一方，データ量・計算量の制限からコードブックサイズはある程度以上大きくできない．そのため，ベクトル量子化の際の量子化誤差が無視できなくなる．また，特徴量にはコードヒストグラムに表される統計量以外にもまだ多くの情報が含まれていると考えられ，それらを効率的に表現する方法が求められている．

そこで，近年，確率・統計理論を用いた頑健なアプローチが考案されている．例えば，2011年に意味インデクシングタスクで1位となったTokyoTech_Canonの手法[5],[6]は，音声を用いた話者識別で発展してきた混合ガウス分布（GMM：Gaussian Mixture Model）を用いる手法を，意味インデクシングに応用したもので，BoWに基づく枠組みを確率論に基づく枠組みへと自然に拡張している．この手法の結果を図1，2に全体の結果と併せて載せてある．

4. マルチメディアイベント検出

4.1 タスクの概要

マルチメディアイベント検出のタスクは2010年度から開始されたタスクである．今年で3年目となるが，まだ始まったばかりのタスクといえよう．1.で述べたように，意味インデクシングが「単語」をクエリとした検索とすれば，このタスクは「文」をクエリとした検索と捉えることができる．すなわち，検索の対象は高次特徴の組合せで構成されることが想定されている．そこでは，高次特徴の選択，複数高次特徴の組合せ法，時系列としての映像の扱いが研究の焦点となることが予想される．

データベースとしてLDC（Linguistic Data Consortium）が提供するHAVICデータベースが用いられる．このデータベースはやはりインターネット映像を集めたものである．2011年は10個のイベントが用意された（表2）．図4にその例を示す．各々のイベントについて，それを含む映像クリップが80〜230個ほど用意される．結果の提出方法，評価基準などは，意味インデクシングとほぼ同様である．2011年は18チームが参加した．

全体として，事前の予想よりも検出性能が高かった．将来の実用化が十分に視野に入っているタスクといえる．参考までに3位となったTokyoTech_Canonの結果[7]を図5に示す．そこでは3.2.5で紹介した枠組みをこのタスクに応用している．

2012年から，提出締切りの直前に検出すべきイベントが公開されるアドホックタスクが追加された．このタスクでは個々のイベントの特徴に依存した識別器を構築することはできず，多くの高次特徴に対する意味インデ

表2 2011年マルチメディアイベント検出タスクで検出対象のイベント

Birthday party	Making a sandwich
Changing a vehicle tire	Parade
Flash mob gathering	Parkour
Getting a vehicle unstuck	Repairing an appliance
Grooming an animal	Working on a sewing project

図4　イベント“Making a sandwich”を含む映像クリップの一例　　サムネイルを時間順に左から右に並べて表示.

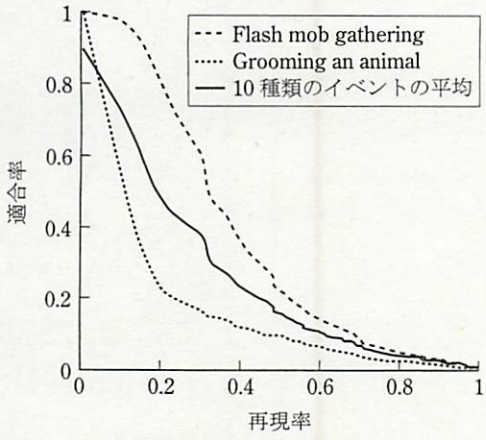

図5　2011年マルチメディアイベント検出タスクにおいて3位となったTokyoTech_Canon[7]の結果　　最もF値が高いイベント(Flash mob gathering)，最もF値の低いイベント（Grooming an animal)，及び，10種類のイベントの平均の三つについて，そのRecall-Precisionカーブを示す.

クシングの結果をどのように使うかが焦点となる.

4.2　検出フレームワーク

現在主流のフレームワークは，映像ショットに対する意味インデクシングのフレームワークを映像クリップに対してそのまま適用するものである．すなわち，個々のイベントに対して，そのイベントを含む映像クリップを正例，それ以外を負例とし，BoWの枠組みに基づく識別器を構築する手法である．特徴量としては，意味インデクシングで用いられている局所特徴に加え，STIP (Space-Time Interest Points) オプティカルフローなどの時空間特徴を表現した特徴量が用いられる.

前節で述べたように，タスクを設定した側は意味インデクシングで得られた高次特徴を活用した手法の開発を意図していた．例えば，各高次特徴に対する識別器のスコアを高次特徴全体について並べたベクトルを特徴量として用いる手法が提案されている．しかし，現時点ではその種の手法は上述のフレームワークに基づく手法に及ばない．また，高次特徴やその組合せの時系列特徴を表現するモデルは実現していない．これらの理由としては，まだタスクが比較的小規模であるため各イベントに特化した識別器の構築が工数的に十分可能であること，現在使用可能な高次特徴の種類が各イベントを記述するためにはまだ不足していること，が挙げられる.

このような状況の中で，将来有望と思われる試みも幾つか見られる．例えば，IBMとコロンビア大のチームが提案しているセマンティックモデル[8]は，Pen Treebank, WordNetなどの自然言語の階層構造（分類木）にならい，LSCOMという高次特徴の大規模な体系を構成し，それをイベント検出に利用しようという試みである．このタスク向けに780種類の画像特徴（「器具」，「釣り道具」，「道具箱」など）の分類木を作成し，その階層構造を利用して画像特徴ごとの検出器を構築している．また，55種類の音響特徴（「人ごみ」，「動物」など），134種類の動作特徴（「蹴る」，「飲む」など）を用意している.

なお，2012年から，MER (Multimedia Event Recounting) タスクが始まった．このタスクは，MEDタスクから更に一歩踏み込んで，イベントを検出した後にその根拠・理由を再説明する（Recounting）タスクである．含まれているイベントが既知の映像クリップが与えられ，参加者はイベントを構成する要素（高次特徴など）の出現時刻や画面上の位置などの検出の根拠・理由となる情報を提出する．主催者側は，提出された情報から，そのクリップに含まれているイベントが何かが分かるか，同じイベントを含む複数の映像クリップの間の違いが分かるか，の2点について主観的に判断を行う．従来の低次特徴のみを用いた手法では，人間の判断に使えるレベルの情報を提示できない．より高次の情報を用いた検出の研究開発を促進するために用意されたタスクである.

5. 今後の展望

ここまで，映像検索・評価ワークショップTRECVIDの意味インデクシング，マルチメディアイベント検出タスクにおける最新技術について解説した.

5.1　特徴抽出とパターンマッチングの境界

パターン認識では特徴抽出の段階を経てパターンマッチングが行われている．特徴抽出の結果得られる特徴量はなるべくコンパクトなものが望ましい．これは主に計算量の制限からくるものである.

例えば，音声認識の音響モデルの分野では，長らくMFCC特徴量とHMMの組合せが使われてきたが，近

年の計算パワーの増加により，プリミティブな極めて高次元の特徴と，それを扱うことのできるより大規模なモデルの組合せにより，性能がより向上することが分かってきた．映像分野においても同様の発展が視野に入ってきている．

5.2 コミュニケーションとしての映像

映像は送り手と受け手が存在する「コミュニケーション」であり，そのコンテンツは送り手の意図と受け手の要求の両方を反映した「言語」であると捉えることができる．そして，その統語規則・意味規則を明示的，あるいは，非明示的に制約として利用することで，映像中の「言語」をより精度良く抽出することが可能になる．

例えば，音声認識では音響信号のモデル（音響モデル）以外に，言語制約としての言語モデルの使用がその実用化に大きく貢献した．それに対し，映像分野では効果的な言語モデルはまだ存在していない．今後の発展が楽しみな分野である．

5.3 高速化

2011年の意味インデクシングタスクでは，参加チームの半数がタスクを完遂できなかった．これは，主に計算リソースの制限から，提出期限までに大量の映像データを処理できなかったためと推測される．好成績を上げたTokyoTech_Canonにおいても，東工大スーパコンピュータTSUBAMEを利用できたことが大きな助けとなった．今後は，他の計算機分野と同様に，並列処理アルゴリズム，GPUの使用が当たり前となるとともに，高速な映像検索アルゴリズムの開発が盛んになるであろう．

謝辞 本稿をまとめるにあたり御議論頂いた東工大とキヤノンの共同チームTokyoTech_Canonの諸氏，特に井上中順氏と上嶋勇祐氏，及び，日頃からチームの活動を御支援頂くキヤノン株式会社に感謝する．

文　献

（1） TREC Video Retrieval Evaluation, http://trecvid.nist.gov/
（2） 佐藤真一，“映像内容解析におけるTRECVIDの取組み，”信学誌，vol. 91, no. 1, pp. 55-59, Jan. 2008.
（3） A.F. Smeaton, P. Over, and W. Kraaij, “Evaluation campaigns and TRECVid,” In Proc. of ACM Multimedia MIR workshop, pp. 321-330, 2006.
（4） MediaEval Benchmark, http://www.multimediaeval.org/
（5） 井上中順，篠田浩一，“木構造GMMを用いたセマンティックインデクシングの高速化，”信学技報，DE2011-19, PRMU2011-50, pp. 105-110, June 2011.
（6） N. Inoue and K. Shinoda, “A fast and accurate video semantic-indexing system using fast MAP adaptation and GMM supervectors,” IEEE Trans. Multimed., vol. 14, no. 4, pp. 1196-1205, 2012.
（7） Y. Kamishima, N. Inoue, K. Shinoda, and S. Sato, “Multimedia event detection using GMM supervectors and SVMs,” Proc. ICIP, 2012 (in press).
（8） L. Cao, S.F. Chang, N. Codella, C. Cotton, D. Ellis, L. Gong, M. Hill, G. Hua, J. Kender, M. Merler, Y. Mu, A. Natsev, and J.R. Smith, “IBM research and columbia university TRECVID-2011 Multimedia Event Detection (MED) system,” Proc. TRECVID2011, 2011.

（平成24年6月4日受付　平成24年6月12日最終受付）

篠田（しのだ）　浩一（こういち）（正員：シニア会員）

1989東大大学院理学系研究科物理学専攻了．同年日本電気株式会社．1997～1998米国ベル研究所客員研究員．2001東大大学院情報理工学系研究科助教授，2003東工大大学院情報理工学研究科准教授．工博．音声・動画像パターン認識の研究に従事．1997年度本会論文賞受賞．